Fall 2023

Complex Microbial Mat Communities Used to Assess Primer Selection for Targeted Amplicon Surveys

Lindsey Smith
Western Washington University, benedil@wwu.edu

Complex Microbial Mat Communities Used to Assess Primer Selection for Targeted Amplicon Surveys

By

Lindsey Smith

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

ADVISORY COMMITTEE

Dr. Craig Moyer, Chair

Dr. Heather Fullerton

Dr. Shawn Arellano

GRADUATE SCHOOL

David L. Patrick, 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.

Lindsey Smith

7 September 2023

Complex Microbial Mat Communities Used to Assess Primer Selection for Targeted Amplicon Surveys

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Lindsey Smith
September 2023

Abstract

The microbiota of hydrothermal vents has been widely implicated in the dynamics of oceanic biogeochemical cycling. Lithotrophic organisms utilize reduced chemicals in the vent effluent for energy, which fuels carbon fixation, and their metabolic byproducts can then support higher trophic levels and high-biomass ecosystems. However, despite the important role these microorganisms play in our oceans, they are difficult to study. Most are resistant to culturing in a lab setting, so culture-independent methods are necessary to examine community composition. Targeted amplicon surveying, in which a marker gene is selected for DNA amplification, has become the standard practice for assessing the structure and diversity of hydrothermal vent microbial communities. The most commonly used marker gene is the small subunit ribosomal RNA (SSU rRNA) gene, due to its ubiquity across all cellular organisms and the presence of both conserved and variable regions. Here, the performance of primer pairs targeting the V3V4 and V4V5 variable regions of the SSU rRNA gene were assessed using environmental samples from microbial mats surrounding iron-dominated hydrothermal vents. Using the amplicon sequence variant (ASV) approach to taxonomic identification, the structure and diversity of microbial communities at Kamaʻehuakanaloa Seamount were elucidated in detail. Both primer pairs generated robust data and comparable alpha diversity profiles. However, several distinct differences in community composition were identified between primer sets, including differential relative abundances of both bacterial and archaeal phyla. The primer choice was determined to be a significant driver of variation among the taxonomic profiles generated. Based on the higher quality of the raw sequences generated and on the breadth of abundant taxa found using the V4V5 primer set, it is determined as the most efficacious primer pair for whole-community surveys of microbial mats at iron-dominated hydrothermal vents.

Acknowledgements

I'm deeply grateful for the guidance and insight of my advisor, Dr. Craig Moyer, in untangling the complex ecology of microbial mats. I would also like to thank Dr. Heather Fullerton for her many, many invaluable contributions to my understanding of bioinformatics. Thank you to Dr. Shawn Arellano for her help in navigating the complexities of the graduate school machine. Much appreciation to my steadfast lab mates, Laura Murray, Emily Musenbrock Aker, Lilja Strang, Amanda Stromecki, and Alyssa Tsukada for their feedback and reassurance. Special thanks to the crew and science team on the R/V Thomas G. Thompson and the Jason pilots, without whom I would have had no data to work on. Finally, I would like to thank my husband, Allen; without his support, encouragement, and computer-wrangling this work would not have been possible.

Table of Contents

List of Tables and Figures

Introduction

Targeted amplicon surveys of the small subunit ribosomal RNA gene (SSU rRNA gene) are a fast and cost-effective culture-independent approach for microbial ecology studies. Since the SSU rRNA gene contains both variable and conserved regions it allows for robust taxonomic identification (McNichol *et al.*, 2021). Additionally, this approach provides relative abundance data that distinguishes both rare and abundant taxa, which can be used to build a picture of community structure and diversity (Pascoal, Costa and Magalhães, 2021). Researchers must choose only short sections of the SSU gene to target and amplify for their analyses. The choice of PCR primers may result in over- or underrepresentation of some groups of taxa in the final dataset, and may lead to inaccurate interpretations or conclusions (Baker, Smith and Cowan, 2003; Hamady and Knight, 2009; Wang and Qian, 2009; Walters *et al.*, 2015; Bahram *et al.*, 2019; Abellan-Schneyder *et al.*, 2021). The design and selection of optimal SSU PCR primers has been an ongoing concern and challenge.

Two of the most commonly used primer sets target either the V3V4 region or V4V5 region of the SSU rRNA gene. A suite of primers was evaluated *in silico* for amplification efficiency and 341F/785R, targeting the V3V4 region, was recommended as the most efficacious pair despite a noted tendency for overrepresentation of Alphaproteobacteria and underrepresentation of both Bacteroidetes and the important SAR11 clade of pelagic bacteria (e.g., order Pelagibacterales*;* Klindworth *et al*., 2013). In 2016, the same strong bias against Pelagibacterales was noted when using the primers adopted by a large-scale sequencing project, the Earth Microbiome Project (EMP), as well as the underrepresentation of two archaeal phyla (Parada, Needham and Fuhrman, 2016). The EMP primers (515F/806R) only amplify the V4 region of the SSU gene, as they were designed at a time when Illumina amplicon length was constrained to 2 x 75-100 bp (Caporaso *et*

al., 2011). To increase amplification efficiency in both Bacteria and Archaea, a different previously-validated reverse primer was selected, and a new degeneracy was added to the forward primer (Quince *et al.*, 2011; Parada, Needham and Fuhrman, 2016). The new primers, referred to as 515F-Y/926R, amplify the V4V5 region of the SSU rRNA gene. Testing against marine samples, the V4V5 primers strongly amplified Pelagibacterales and recovered multiple additional taxa not covered by the V4 primers. The V4V5 primers were tested against mock communities and produced community profiles that much more closely matched the expected composition than those generated by the V4 primers in both even and staggered trials (Parada, Needham and Fuhrman, 2016).

Mock communities are small synthetic constructs of curated taxa (typically 10-20 representatives) with known clonal abundances, so they generate quantifiable outcomes in high-throughput sequencing comparison studies (Shakya *et al.*, 2013). Studies have shown the importance of comparison to environmental samples rather than simply relying on mock community data when evaluating the accuracy of primer performance (Wear *et al.*, 2018; Abellan-Schneyder *et al.* 2021). This was demonstrated during the development of the V4V5 primers: the differential abundance of Pelagibacterales detected by V4V5 was less than twofold that detected by the V4 primers in mock analysis, while in the environmental samples it was between 4- and 10-fold higher (Parada, Needham and Fuhrman, 2016). Established microbial mat communities can be highly complex (Vander Roost, Thorseth and Dahle, 2017), and with increasing community complexity challenges exist for evaluating differences in community composition between primers because the community composition is unknown. A comparison of primer performance between several mock communities, saliva samples, and soil samples indicated that the degree of difference between community profiles generated by different primers may scale with increasing community

complexity (Soriano-Lerma *et al.*, 2020). For this reason, the assessment of primers for use on highly complex communities, like those found in microbial mats, should be conducted directly on environmental samples.

Hydrothermal vent systems represent a gradient at the interface between the oceanic crust and the deep-sea water column, presenting an ideal habitat for lithotrophic microorganisms driven by an abundance of dissolved reduced compounds. Iron, the most abundant element in the Earth, is a limiting micronutrient for marine phototrophic primary production (Boyd *et al.*, 2007). Ferrous iron is rapidly oxidized abiotically, drawing it out of solution and dramatically reducing its bioavailability. However, analysis of plumes from high temperature focused venting shows that more iron than expected stays dissolved in the water column, and can be transported in the plume for great distances into the upper ocean (Neuholz *et al.*, 2020; Resing *et al.*, 2015). Microbes at the vent orifices may play a direct role in iron transport and iron speciation (H. Wang *et al.*, 2021). Recent studies support these hypotheses in the dynamics of iron transport at a diffuse vent system as well, and when comparing iron speciation from different types of vents, diffuse vent flow hosts the highest proportion of labile iron complexed with organic material (Lough *et al.*, 2019; W. Wang *et al.*, 2021). Therefore, diffuse vent systems that host microbial mats and release fluid high in dissolved iron, like those observed at Kama‘ehuakanaloa Seamount, are of particular importance (Wheat *et al.*, 2000; Glazer and Rouxel, 2009). The vast majority of the organisms living in the microbial mats surrounding the vents are not yet culturable, so culture-independent methods are necessary to study these communities and elucidate their impact on global biogeochemical cycles (Emerson and Moyer, 2010; Duchinski *et al.*, 2019).

Researchers commonly choose either the V3V4 or V4V5 region to amplify in hydrothermal vent-associated microbial mat diversity surveys (Hager *et al.*, 2017; Scott, Glazer and Emerson,

2017; Duchinski *et al.*, 2019; Ramírez *et al.*, 2021; Speth *et al.*, 2022; Stromecki *et al.*, 2022; Astorch-Cardona *et al.*, 2023). In this study, primer pairs for the V3V4 region (Klindworth *et al.*, 2013) and the V4V5 region (Parada, Needham and Fuhrman, 2016), were used to amplify the SSU rRNA gene from microbial mats around the iron-dominated diffuse hydrothermal vents at Kama‘ehuakanaloa Seamount to determine if significant differences in community composition or structure would result. A deep sequencing effort of these complex communities was undertaken for both primer sets. Upon analysis, several distinct differences in composition between datasets were revealed in both the Bacterial and Archaeal domains. A comparison of overall taxonomic profiles and relative abundance data reveals that both primer sets generate robust data, but that primer choice is significant in differences among the datasets associated with iron-dominated microbial mat sequencing.

Methods

Sample Collection

Samples were collected from Kamaʻehuakanaloa Seamount (formerly known as Lōʻihi Seamount), an active undersea volcano at the leading edge of the Hawaiian hot spot (Clague *et al.*, 2019). Four locations around the seamount were selected for sampling: Pohaku, at the southernmost edge of the caldera; Lohiau, on a small outcrop on the northern lip of the caldera; Hiolo North, on the eastern lip of the caldera; and Hiolo South, just south of Hiolo North (Table 1). At each of the four locations, two suction samples were collected. All venting locations emitted fluid that ranged from 20-48ºC and all were surrounded by abundant microbial mat (Fullerton *et al.*, 2017; Scott, Glazer and Emerson, 2017).

Microbial mats were collected in March 2013 on the R/V *Thomas G. Thompson* cruise TN293 with ROV *Jason II*. Eight bulk mat samples were collected with an impeller-driven suction device into chambers using a double layer 202 µm Nytex catchment barrier. Samples were homogenized and then aseptically transferred to sterile 50ml centrifuge tubes inside a cold room, preserved in RNA*later* (ThermoFisher Scientific, Waltham, MA), and stored at -80°C until DNA extractions.

DNA extraction, Amplification, and Sequencing

Genomic DNA was extracted from each sample in triplicate using the FastDNA SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA) after fluid removal, as previously described (Jesser *et al.*, 2015; Hager *et al.*, 2017). Extracted DNA was quantified using a Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA).

Each sample collection was PCR amplified in quintuplicate using each primer set, for sixteen pooled amplicon sets across the eight sample collections. The V3V4 region of the SSU

rRNA gene was amplified using 341F (5'-CCTACGGGNGGCWGCAG-3') and 785R (5'-GGACTACHVGGGTATCTAATCC-3'), and the V4V5 region was amplified using 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 926R (5'-CCGYCAATTYMTTTRAGTTT-3'), then amplicons were cleaned and indexed according to Illumina MiSeq best-practices (Illumina, 2013). Amplicon libraries were quantified with a Qubit 2.0 fluorometer. Sequencing was performed at Western Washington University on an Illumina MiSeq using manufacturer protocol (Illumina, Inc.), generating 2 x 300 bp paired-end reads.

Sequence Processing

Raw FASTQ files were first primer-trimmed with Cutadapt (Martin, 2011) using the linked-adapter protocol with the complete primer sequences; untrimmed sequences were discarded. Trimmed FASTQ files were then imported to R and each primer set was run independently using the DADA2 analysis pipeline, following established best practices as previously described (Callahan *et al.*, 2016). After evaluation of the read quality plots, ideal positions for truncating the reverse primers were determined to be at the 20-quality score threshold, based on maximal retention of subsequent merged reads. The core inference algorithm was applied using pseudopooling. Taxonomy was assigned to the genus level using the DECIPHER R package with the IDTAXA command (Quast *et al.*, 2013; Murali, Bhargava and Wright, 2018), aligning to the SILVA reference database v138 (Yilmaz *et al.*, 2014). Finally, ASVs assigned to the class Zetaproteobacteria were analyzed by ZetaHunter (https://github.com/mooreryan/ZetaHunter) for classification to previously defined Zetaproteobacterial OTUs (e.g. zOTUs; McAllister, Moore and Chan, 2018).

ASVs from each primer pair that were unidentified at the domain level were removed from analysis. To combine the two datasets, ASVs identified to at least the domain level were then

iteratively clustered down to the genus level such that ASVs binned inside a single genus were condensed into a single entry in the taxonomic table, ASVs that were unclassified at the genus level were clustered by family, and so on. Abundances of each ASV in each cluster were also condensed per sample. This compressed taxonomy is subsequently referred to as the "condensed taxa" and allowed for direct comparisons between primer sets.

Diversity and Statistical Analysis

Rarefactions were plotted for the non-condensed ASV data per primer set using tidyverse (Wickham *et al.*, 2019). Alpha diversity metrics (Chao1 Richness, Shannon Diversity, Simpson Evenness) were calculated for non-condensed ASV data in each sample using the microbiome package (Lahti and Shetty, 2012). Prior to analysis, data in each metric were checked for normal distribution using the Shapiro-Wilke test. Differences between primer set were assessed for statistical significance in each metric for the combined sample data: Analysis of Variance was used for Shannon Diversity and Chao1 Richness and Kruskal-Wallis was used for Simpson Evenness.

The condensed taxa sample data matrices were read into R using phyloseq (McMurdie and Holmes, 2013) and then normalized using variance stabilizing transformation (VST) with package DESeq2 (Love, Huber and Anders, 2014). Plots were drawn using ggplot2 (Wickham, 2016). Since VST-normalized data contain negative values, the Euclidean distance matrix was calculated for the normalized data. Normalized data were PCoA plotted. The vegan package was used to analyze the distance matrix; first, beta dispersion by primer choice was validated as non-significant using betadisper2, then PERMANOVA via adonis2 was performed using primer choice as the variable (Oksanen *et al.,* 2022). PERMANOVA to assess location as a driver of variation between primer sets could not be conducted because the beta dispersion homogeneity assumption was not met for that variable.

Results

Sequence Processing

The chosen sequencing approach resulted in sixteen community profiles representing eight collections each, divided by the two primer pairs. Amplicon processing through the DADA2 pipeline generated a total of 7,506 ASVs in 1,952,073 reads from the eight communities amplified by the V3V4 primers. This output represents a net loss of 63.08% of the total input reads, with 54.61% to 75.11% of reads lost from a single community (Supplemental Table 1). After removing ASVs that were unclassified at the domain level, the V3V4 dataset comprised 4,140 ASVs in 1,558,201 reads, meaning that 44.8% of sequences could not be identified to at least the domain level. DADA2 processing of the eight communities amplified by the V4V5 primers generated a total of 5,130 ASVs in 2,950,958 reads. This translates to a net loss of 52.17% of the total input reads, ranging from 48.43% to 57.27% of reads lost from a single community (Supplemental Table 2). After removing ASVs that were unclassified at the domain level, the V4V5 dataset comprised 3,288 ASVs in 2,227,954 reads, meaning that 35.9% of reads could be not identified to at least the domain level.

Community Profiling

In constructing the rarefaction curves for the V3V4 primer set, mean depth was 244,009 reads per community with a minimum of 130,311 and a maximum of 323,532. The V4V5 rarefaction curves had a mean depth of 335,119 reads per community with a minimum of 316,860 and a maximum of 365,477. Plateau was reached in all communities from both primer sets (Supplemental Figure 1).

After calculating alpha diversity metrics, no significant differences between primer sets in Chao1 Richness was found using Kruskal-Wallis and no significant difference in Shannon

Diversity or Simpson Evenness were found using Analysis of Variance (Figure 1). However, PERMANOVA analysis confirmed that primer set was a significant driver of the differences observed among the taxonomic profiles (R^2=0.25, P=0.001). PCoA of the normalized data explained 51.6% of total variance in both principal coordinates and showed distinct clusters associated with each primer set at the 95% confidence interval (Figure 2). No significant clustering was associated with location.

Domain Archaea

The V3V4 primer set generated 62 Archaeal ASVs spread over six phyla plus another group that was unclassified except at the domain level. Twenty-six ASVs were identified as Crenarchaeota, 13 as Thermoplasmatota, eight each as Hydrothermarchaeota and Halobacterota, three as Nanoarchaeota, three were unclassified at the phylum level, and a single ASV identified as Euryarchaeota. Phylum Asgardarchaeota was unrepresented in the V3V4 dataset (Figure 3a). By relative abundance in domain Archaea, Thermoplasmatota was the most abundant phylum with 41.79% of reads, followed by Crenarchaeota with 22.81% and Hydrothermarchaeota with 20.49% (Figure 3b).

The V4V5 primer set generated approximately two-thirds again as many Archaeal ASVs as the V3V4 primer set, with 103 ASVs spread over six phyla plus another group that was unclassified except at the domain level. Crenarchaeota was the dominant phylum in terms of both number of representative ASVs (53) and in relative abundance within domain Archaea (63.42% of archaeal reads; Figures 3a and 3c). Phylum Thermoplasmatota was represented by 19 ASVs, 11 were classified as Nanoarchaeota (the next greatest by relative abundance, with 10.26% of reads), seven as Hydrothermarchaeota, four as Halobacterota, and three unclassified at the phylum level. All but Crenarchaeota and Nanoarchaeota composed less than 10% of total archaeal reads (Figure

3c). In contrast to the V3V4 samples, phylum Asgardarchaeota was represented by seven ASVs and phylum Euryarchaeota was unrepresented (Figure 3a).

Of the eight identified divisions in domain Archaea (inclusive of unclassified phyla), six differed by at least one order of magnitude in relative abundance between primer sets. Halobacterota and Crenarchaeota abundances were within the same order of magnitude between primers, although V4V5 had more than three times as many reads identified as Crenarchaeota than did V3V4. While both primers detected one phylum unique to each dataset, phylum Euryarchaeota in V3V4 was represented by a single ASV that accounted for 0.06% relative abundance and phylum Asgardarchaeota in V4V5 was represented by 7 ASVs that accounted for 0.95% relative abundance (Figures 3a, 3b, 3c). Only Halobacterota abundances were similar between primer sets, despite differences in representative ASVs (Figures 3a, 3b, 3c).

Domain Bacteria

There were 4,078 ASVs recovered from the V3V4 primer set that classified to domain Bacteria, belonging to 49 phyla (Figure 4a, some data not shown). Of those, 85.1% (3,471 ASVs) belonged to phyla that represented greater than 1% relative abundance, and a further 527 ASVs belonged to phyla that represented less than 1% but greater than 0.1% relative abundance. The phyla with the greatest number of ASVs were Proteobacteria with 1,322, followed by Planctomycetota (498 ASVs), Bacteroidota (262), Patescibacteria (sometimes called Candidate Phyla Radiation, or CPR; 206), Verrucomicrobiota (146), and Chloroflexi (135). There were 709 ASVs that remained unclassified below the domain level. By relative abundance, 12 phyla made up at least 1% of the reads, respectively (Figure 4b). Proteobacteria made up fully half of the bacterial reads (50.73%), with the next most abundant phyla being Planctomycetota (8.11%), Bacteroidota (7.17%), Patescibacteria (6.32%), and the uncharacterized DTB120 (4.83%). Phyla

with relative abundance between 0.99% and 0.1%, were led by Gemmatimonadota, Acidobacteriota, Zixibacteria, Verrucomicrobiota, and Bdellovibrionota. The lowest relative abundance phyla (those under 0.1% abundance) were the most numerous, with 22 different representatives, but their 77 ASVs altogether accounted for just 0.28% of total bacterial reads.

The V4V5 primer set generated 3,185 ASVs classified to domain Bacteria, representing 55 phyla (Figure 4a, some data not shown). Phyla that made up a least 1% relative abundance, respectively, accounted for 2,443 of those ASVs (76.7% of bacterial ASVs), while phyla between 0.99% and 0.1% relative abundance accounted for a further 629 ASVs. The phyla with the most representative ASVs were Proteobacteria with 979, followed by Planctomycetota (447 ASVs), Bacteroidota (289), Chloroflexi (128), and Verrucomicrobiota (121). There were 393 ASVs unclassified below domain level. By relative abundance, eight phyla contributed greater than 1%, respectively (Figure 4c). Proteobacteria made up more than half of all reads (58.44%), followed by Bacteroidota (9.80%), Planctomycetota (5.61%), Nitrospirota (5.42%), and DTB120 (4.77%). The top five most abundant phyla between 0.1% and 1% relative abundance were MBNT15, Gemmatimonadota, Zixibacteria, Patescibacteria, and Bdellovibrionota. There were 28 phyla with relative abundances below 0.1%, whose 103 ASVs comprised 0.37% of total bacterial reads.

There were several dissimilarities between the two bacterial taxonomic profiles. Two phyla were unique to the V3V4 dataset (GAL15 and Poribacteria) and eight phyla were unique to the V4V5 dataset (Caldatribacteriota, CK-2C2-2, Deferrisomatota, Edwardsbacteria, Elusimicrobiota, FCPU426, Latescibacterota, and TA06). Four phyla with greater than 0.1% relative abundance in at least one primer differed in that abundance by at least one order of magnitude: Patescibacteria (V3V4: 6.32%, V4V5: 0.50%), Campylobacterota (V3V4: 2.19%, V4V5: 0.33%), Acetothermia (V3V4: 0.34%, V4V5: 0.03%), and Deinococcota (V3V4: 0.002%, V4V5: 0.16%) (some data not

shown). Patescibacteria and Campylobacterota also differed more than any other taxa in numbers of representative ASVs between primer sets, with V3V4 having 3.5-fold more Patescibacteria ASVs and 3-fold more Campylobacterota ASVs. There were two additional taxonomic groupings with notable variation in ASVs between primers: phylum Proteobacteria and taxa unclassified below domain ranking (Figure 4a). While those two groupings contained 659 ASVs unique to the V3V4 dataset (approximately 75% of the total number of ASVs unique to V3V4), all the unique ASVs were likely present in very low abundance (Figures 4b and 4c). However, four out of the top five most abundant phyla in each primer set were the same and their numbers of representative ASVs were within the same order of magnitude.

Phyla Proteobacteria and Campylobacterota

The V3V4 primer set generated 1,322 ASVs identified to phylum Proteobacteria and 83 phylum Campylobacterota ASVs. Within the two phyla, class Zetaproteobacteria was represented by 266 ASVs and made up 61.18% of total reads (Figures 5a and 5b). In contrast, class Gammaproteobacteria was the most diverse with 623 ASVs, but accounted for just 21.39% of total reads. Next greatest by abundance was class Alphaproteobacteria with 350 ASVs constituting 13.14% of reads, followed by a mere 0.16% of reads consisting of 83 ASVs unclassified below phylum level in Proteobacteria (Figure 5a). All Campylobacterota ASVs were also classified as class Campylobacteria and composed 4.13% of reads.

Phylum Proteobacteria had 979 ASVs in the V4V5 dataset and phylum Campylobacterota had just 27 ASVs (all of which were also identified as class Campylobacteria). Class Zetaproteobacteria repeated the trend observed in the V3V4 dataset but to an even greater extent: by relative abundance, Zetaproteobacteria were in the definitive majority with 74.98% of total reads but were represented by only 89 ASVs (Figures 5a and 5c). Gammaproteobacteria profiles

were also similar to V3V4, with an extensive collection of 531 ASVs comprising only 15.54% of total reads within the two phyla. Class Alphaproteobacteria had nearly the same number of ASVs as were present in the V3V4 dataset but they made up only about two-thirds as much of the V3V4 relative abundance. Only 26 ASVs remained unclassified below the phylum level in Proteobacteria and they accounted for 0.35% of total reads. Class Campylobacteria made up only 0.56% of total reads.

Two interesting differences among the taxonomic profiles within these data are the relative abundances of both Campylobacteria and Zetaproteobacteria. The relative abundances of class Campylobacteria differed by an order of magnitude, with V3V4 having the greater representation. Zetaproteobacteria abundances differed to a lesser degree, with V4V5 containing approximately 15% relatively more Zetaproteobacteria reads. The relative discrepancy of Zetaproteobacteria reads at the class level translated to a difference of more than 10% in the relative abundance of Zetaproteobacteria between the two datasets as a whole; Zetaproteobacteria accounted for 32.28% of total identified reads in V3V4 and 45.54% of total identified reads in V4V5.

Class Zetaproteobacteria

There were 26 previously-established Zetaproteobacteria OTUs (zOTUs) found with the V3V4 primer set (Figure 6a). By far the most abundant was zOTU2, with 56.46% of the total zOTU reads (Figure 6b). zOTU1 followed with 15.76%, then zOTU4 (8.20%), zOTU10 (7.11%), and zOTU14 (2.09%). The majority of the remaining zOTUs were represented by less than 1% of the zOTU reads (Figure 6c). Additionally, 75 sequences were classified as “NewZetaOTUs”, all of which were found in very low abundance (fewer than 20 total reads each).

The V4V5 primers retrieved 24 previously-established zOTUs (Figure 6a). As with the V3V4 dataset, zOTU2 was the most dominant taxon with 55.99% of the zOTU reads (Figure 6d).

zOTU1 was next in abundance with 9.45%, then zOTU14 (9.56%), zOTU10 (6.41%), and zOTU4 (4.63%). Similar to the V3V4 dataset, the majority of identified zOTUs were present in less than 1% relative abundance (Figure 6e). Only 16 "NewZetaOTUs" were found by the V4V5 primers, all in very low abundance.

The V4V5 primers retrieved nearly twice as many total Zetaproteobacteria reads as did V3V4 (990,865 in V4V5 compared to 502,938 in V3V4). Both V3V4 and V4V5 identified the same eight zOTUs as most abundant and although the absolute number of reads in each zOTU, respectively, differed by up to nine-fold (zOTU14, Figure 6a), when comparing the data by relative abundance the profiles are quite similar. This trend in absolute versus relative abundance is also observed in the lower abundance zOTUs (Figures 6a, 6c, and 6e). However, there are several exceptions to the similarities in the lower abundance zOTUs. Two zOTUs differed by at least one order of magnitude in relative abundance: zOTU18 (V3V4: 0.074%, V4V5: 1.047%) and zOTU26 (V3V4: 0.029%, V4V5: 0.001%). Additionally, there were zOTUs present in one dataset that were completely absent in the other: zOTU54 and zOTU12 were not found using the V3V4 primers, while zOTUs 52, 31, 23, and 40 were missing from the V4V5 data. All unique zOTUs, regardless of dataset, were represented by less than 1% relative abundance, and all but two (zOTU52 in V3V4 and zOTU54 in V4V5) were represented by less than 0.1% relative abundance (Figures 6d and 6e).

Discussion

Primer selection has previously been evaluated as a source of potential bias in microbial ecology studies of various complex communities. Primer selection has been determined to be significant as a driver of variation among the datasets generated from studies on the human microbiome, soils, terrestrial mineral deposits, the pelagic water column, marine subsurface sediment, and sea ice (Walters *et al.*, 2015; Pollock *et al.*, 2018; Bahram *et al.*, 2019; Rajeev *et al.*, 2020; Fadeev *et al.*, 2021; Hathaway *et al.*, 2021; McNichol *et al.*, 2021; Nearing, Comeau and Langille, 2021). Targeted amplicon surveys of iron-dominated microbial mats have been an important analytical tool for assessing diversity and community structure, but as yet there is no consensus on the optimal primer pair to use for such surveys. The V4V5 primers used here show greater promise than the V3V4 primers for analysis of hydrothermal vent-associated iron-dominated microbial mats, and these results have implications for the study of other complex microbial communities.

The protocols used in this study were chosen after careful review of current methodological literature. A recent study undertook a thorough comparison of various ASV amplicon pipeline step-wise processes and highlighted both the downstream influence of forward and reverse sequence truncation before merging and the impact that reference database has on taxonomic identification (Abellan-Schneyder *et al.*, 2021). Based on that research, forward and reverse reads for this study were custom trimmed after repeated testing to remove low-quality bases while maintaining maximal retention of subsequent merged reads for further processing. The per-base quality of the input sequences varied between primer sets, with the V3V4 reverse sequences showing an abrupt drop in quality around position 250. That drop in quality was not seen in the corresponding V4V5 reverse sequences. The low-quality bases in the V3V4 sequences

necessitated much more liberal trimming to maintain acceptable quality scores. This V3V4 trimming may be a situation where "the baby was thrown out with the bathwater"; effectively, because the V4V5 sequences were of higher initial quality, more material remained to be analyzed by the pipeline. This concept is expanded when considering the ratios of completely unclassified taxa between the two primer sets; while the V3V4 primers produced more ASVs, a greater proportion were unidentifiable than those in the V4V5 dataset. While it is possible that improvements to available reference databases may ameliorate this discrepancy in the future, the use of the up-to-date SILVA curated database for taxonomic assignment ensured as comprehensive a community profile as possible at this time (Abellan-Schneyder *et al.*, 2021). Sequencing depth has been shown to influence measures of alpha diversity in microbial ecology studies, both in terms of richness and evenness (Reese and Dunn, 2018; Ramakodi, 2021; Kleine Bardenhorst *et al.*, 2022). In undertaking a deep sequencing effort, biases in alpha diversity measures have been minimized, therefore assuring the robustness of the generated datasets. The lack of statistically significant differences in alpha diversity profiles between primer sets confirms the suitability of downstream comparisons of taxonomic profiles.

Differences in taxonomic profiles were apparent at each ranking examined, which is consistent with the findings of other studies that compared these two primer pairs using environmental samples. Primer choice exhibited a strongly detectible effect in a study of terrestrial soil microbial communities at both the phylum and genus levels, with the difference between V3V4 and V4V5 more pronounced than when comparing V4V5 to other regions (Soriano-Lerma *et al.*, 2020). In that same study, V4V5 was also able to detect variation in composition across the broadest range of community complexity as measured by soil development, compared to other hypervariable regions. However, when they compared the less complex microbial community of

human saliva, the effect of primer choice was weaker. In an examination of various arctic microbial communities, taxa at both the class and family level were differentially abundant based on V3V4 or V4V5 primer choice (Fadeev *et al.*, 2021). The differences among community profiles increased in amplitude with increasing community complexity, with the greatest differential present in the marine sediment community. Interestingly, even though 14% of the total lineages found with V3V4 in that study were absent from the V4V5 profile, those lineages accounted for less than 1% of the total sequences in the V3V4 dataset. This same trend was observed in the current study: while V3V4 produced a greater number of taxa, those taxa that were unique to V3V4 were almost exclusively of very low abundance. The trend is continued when the synthesis of ASVs per taxon and the relative abundance of those taxa are compared between primer sets. Despite the relative abundance data looking similar for each primer set, that abundance is divided into more ASVs in the V3V4 dataset; there are more ASVs sharing each piece of the pie, indicating that there are fewer representatives per taxon than in the V4V5 dataset. Moreover, the relative abundance data are validated as independent of sequencing effort by the plateaus reached in the rarefaction curves, which show that increased sequencing depth would not reveal additional diversity.

The high-resolution taxonomic analysis of class Zetaproteobacteria, a clade of iron-oxidizing bacteria, is of particular importance with regard to the study of biogeochemistry at iron-dominated diffuse hydrothermal vents. An alignment and analysis of the sequences generated by both primer sets should be undertaken to determine the loci of variations that convey taxonomic identity, to determine if the few zOTU holes in the V4V5 dataset can be filled. As part of their iron metabolism, Zetaproteobacteria produce extracellular structures composed of Fe(III) waste product complexed with carboxyl-rich polysaccharides (Chan *et al.*, 2011). Iron that has been complexed with organic components remains in the dissolved fraction of total iron in the water

column much more efficiently than iron complexed with inorganic colloids, and spectral and micro-X-ray fluorescence analysis of diffuse plume-associated iron-bearing particles revealed them to be rich in Fe(III), carbon, and oxygen (Lough *et al.*, 2019). This profile is consistent with that of the biomineral stalks produced by Zetaproteobacteria (Chan *et al.*, 2011). Iron complexation with exopolysaccharides has been shown to enhance bioavailability to marine phytoplankton, which have been estimated to contribute approximately half of planetary primary production (Field *et al.*, 1998; Hassler *et al.*, 2011, 2015). The ability to accurately catalog the Zetaproteobacterial diversity at diffuse vents is a necessary step in the evaluation of how microbial activity in these systems may be shaping primary productivity in the surrounding oceans.

When assessing experimental methodologies, the question at hand must always be the greatest consideration. Across the breadth of taxonomic rankings, the V4V5 primer set tested here revealed the greatest range of the most abundant taxa present in the study system, particularly due to the inclusion of a much broader assortment of domain Archaea. The V3V4 primer set presented a slightly different picture, with often significantly fewer representatives of a larger overall community. However, caveats exist for both primer sets. The underrepresentation of archaea is quite a large hindrance in the use of V3V4 to survey microbial communities as a whole. While examining performance of primers targeting archaea specifically, the V3V4 region has been shown to have a high incidence of non-specific amplification (Fischer *et al.*, 2016); the greater loss of reads through processing observed here is consistent with those results. Conversely, the underrepresentation of the class Campylobacteria in the V4V5 dataset may be problematic for some research questions, particularly in ecosystems rich in sulfur (Dick, 2019). A recent study used the V3V4 primers in an examination of the potential effects of climate change on the composition and ecological function of salt marsh microbial mats; the analysis presented here may

suggest a reassessment of the completeness of their evaluations (Mazière *et al.*, 2023). The impact of environmental factors on hydrothermal microbial community development in the sulfur-rich Guaymas Basin field was recently assessed using V4V5 primers (Ramírez *et al.*, 2021); the current study suggests that the use of those primers might limit their findings.

Choosing PCR primers for environmental surveys can be difficult. While employing multiple sequencing efforts using different primers to capture the diversity of different taxa may be practicable, the cost-effectiveness of using a single primer pair to capture as much diversity as possible is desirable. The V4V5 primers detected a greater number of archaeal taxa and a greater range of bacterial phyla. Given that the V4V5 primers were able to amplify some of the underrepresented taxa rather than missing them entirely, reevaluating the existing 515F-Y and 926R primers for possible additional degeneracies seems the more plausible scenario for bridging the taxonomic gaps between the two primer sets. Particular attention should be given to determining the cause of the "blind spots" within the spectrum of zOTUs captured by V4V5. A novel protocol was recently developed for validating primer combinations against reference databases to select oligonucleotide mixtures representing only variants identified within the reference database or in the natural environment, preserving PCR efficiency as much as possible (McNichol *et al.*, 2021). Additionally, using a combination of metagenomic and amplicon approaches, several loci in the V4V5 forward primer have already been identified where added degeneracies would resolve the majority of missing Patescibacteria, the one exception to the shared top five most abundant taxa between the two primer sets (McNichol *et al.*, 2021). Work to improve the V4V5 primers evaluated here should continue along this trajectory.

Conclusions

Based on the initial higher quality of the raw sequences and the composition of the taxonomic profiles produced by this study, the V4V5 region of the SSU rRNA gene is the more efficacious area for targeted amplicon surveys of iron-dominated microbial mats. However, the results of this study apply particularly to this study system. The high specificity of microbiota per environment may preclude direct between-environment comparisons of primer performance. It is only through repeated per-environment examinations of primer performance that sturdy foundations for future studies can be built. Furthermore, sequence analysis is only as good as the reference databases used, so continued expansion of high-quality curated reference sequence databases is crucial. As reference databases improve, PCR primer design will improve. As primer design improves, not only do the established amplicon protocols improve, but developing methodologies like hybrid amplicon-shotgun metagenomic surveys gain greater rigor.

Works Cited

Abellan-Schneyder, I. *et al.* (2021) 'Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing', *mSphere*, 6(1), pp. 1–22. Available at: https://doi.org/10.1128/msphere.01202-20.

Astorch-Cardona, A. *et al.* (2023) 'Spatial comparison and temporal evolution of two marine iron-rich microbial mats from the Lucky Strike Hydrothermal Field, related to environmental variations', *Frontiers in Marine Science*, 10, p. 1038192. Available at: https://doi.org/10.3389/fmars.2023.1038192.

Bahram, M. *et al.* (2019) 'Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment', *Environmental Microbiology Reports*, 11(4), pp. 487–494. Available at: https://doi.org/10.1111/1758-2229.12684.

Baker, G.C., Smith, J.J. and Cowan, D.A. (2003) 'Review and re-analysis of domain-specific 16S primers', *Journal of Microbiological Methods*, 55(3), pp. 541–555. Available at: https://doi.org/10.1016/j.mimet.2003.08.009.

Boyd, P.W. *et al.* (2007) 'Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions', *Science*. Available at: https://doi.org/10.1126/science.1131669.

Callahan, B.J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. Available at: https://doi.org/10.1038/nmeth.3869.

Caporaso, J.G. *et al.* (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the National Academy of Sciences of the United States of America*, 108(SUPPL. 1), pp. 4516–4522. Available at: https://doi.org/10.1073/pnas.1000080107.

Chan, C.S. *et al.* (2011) 'Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation', *The ISME Journal*, 5(4), pp. 717–727. Available at: https://doi.org/10.1038/ismej.2010.173.

Clague, D.A. *et al.* (2019) 'Structure of Lō'ihi Seamount, Hawai'i and Lava Flow Morphology From High-Resolution Mapping', *Frontiers in Earth Science*, 7, p. 58. Available at: https://doi.org/10.3389/feart.2019.00058.

Dick, G.J. (2019) 'The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally', *Nature Reviews Microbiology*, 17(5), pp. 271–283. Available at: https://doi.org/10.1038/s41579-019-0160-2.

Duchinski, K. *et al.* (2019) 'Fine-Scale Biogeography and the Inference of Ecological Interactions Among Neutrophilic Iron-Oxidizing Zetaproteobacteria as Determined by a Rule-Based Microbial Network', *Frontiers in Microbiology*, 10, p. 2389. Available at: https://doi.org/10.3389/fmicb.2019.02389.

Emerson, D. and Moyer, C. (2010) 'Microbiology of Seamounts: Comon Patterns Observed in Community Structure', *Oceanography*, 23(1), pp. 148–163.

Fadeev, E. *et al.* (2021) 'Comparison of Two 16S rRNA Primers (V3–V4 and V4–V5) for Studies of Arctic Microbial Communities', *Frontiers in Microbiology*, 12, p. 637526. Available at: https://doi.org/10.3389/fmicb.2021.637526.

Field, C.B. *et al.* (1998) 'Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components', *Science*, 281(5374), pp. 237–240. Available at: https://doi.org/10.1126/science.281.5374.237.

Fischer, M.A. *et al.* (2016) 'Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High Reproducibility within and Low Comparability between

Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs', *Frontiers in Microbiology*, 7. Available at: https://doi.org/10.3389/fmicb.2016.01297.
Fullerton, H. *et al.* (2017) 'Hidden diversity revealed by genome-resolved metagenomics of iron-oxidizing microbial mats from Lō'ihi Seamount, Hawai'i', *The ISME Journal*, 11(8), pp. 1900–1914. Available at: https://doi.org/10.1038/ismej.2017.40.
Glazer, B.T. and Rouxel, O.J. (2009) 'Redox speciation and distribution within diverse Iron-dominated microbial habitats at Loihi Seamount', *Geomicrobiology Journal*, 26(8), pp. 606–622. Available at: https://doi.org/10.1080/01490450903263392.
Hager, K.W. *et al.* (2017) 'Community Structure of Lithotrophically-Driven Hydrothermal Microbial Mats from the Mariana Arc and Back-Arc', *Frontiers in Microbiology*, 8, p. 1578. Available at: https://doi.org/10.3389/fmicb.2017.01578.
Hamady, M. and Knight, R. (2009) 'Microbial community profiling for human microbiome projects: Tools, techniques, and challenges', *Genome Research*, 19(7), pp. 1141–1152. Available at: https://doi.org/10.1101/gr.085464.108.
Hassler, C.S. *et al.* (2011) 'Saccharides enhance iron bioavailability to Southern Ocean phytoplankton', *Proceedings of the National Academy of Sciences*, 108(3), pp. 1076–1081. Available at: https://doi.org/10.1073/pnas.1010963108.
Hassler, C.S. *et al.* (2015) 'Iron associated with exopolymeric substances is highly bioavailable to oceanic phytoplankton', *Marine Chemistry*, 173, pp. 136–147. Available at: https://doi.org/10.1016/j.marchem.2014.10.002.
Hathaway, J.J.M. *et al.* (2021) 'A Comparison of Primers in 16S rRNA Gene Surveys of Bacteria and Archaea from Volcanic Caves', *Geomicrobiology Journal*, 38(9), pp. 741–754. Available at: https://doi.org/10.1080/01490451.2021.1943727.
Illumina (2013) '16S metagenomic sequencing library preparation protocol: preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. Part no. 15044223 Rev B.' Illumina, San Diego, CA. Available at: https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
Jesser, K.J. *et al.* (2015) 'Quantitative PCR Analysis of Functional Genes in Iron-Rich Microbial Mats at an Active Hydrothermal Vent System (Lō'ihi Seamount, Hawai'i)', *Applied and Environmental Microbiology*. Edited by G. Voordouw, 81(9), pp. 2976–2984. Available at: https://doi.org/10.1128/AEM.03608-14.
Kleine Bardenhorst, S. *et al.* (2022) 'Richness estimation in microbiome data obtained from denoising pipelines', *Computational and Structural Biotechnology Journal*, 20, pp. 508–520. Available at: https://doi.org/10.1016/j.csbj.2021.12.036.
Klindworth, A. *et al.* (2013) 'Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies', *Nucleic Acids Research*, 41(1), pp. 1–11. Available at: https://doi.org/10.1093/nar/gks808.
Lahti, L and Shetty, S (2012) 'microbiome package for R'.
Lough, A.J.M. *et al.* (2019) 'Diffuse hydrothermal venting: A hidden source of iron to the oceans', *Frontiers in Marine Science*, 6(JUL), pp. 1–14. Available at: https://doi.org/10.3389/fmars.2019.00329.
Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. Available at: https://doi.org/10.1186/s13059-014-0550-8.

Martin, M. (2011) ‘Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads’, *EMBnet Journal*, 17(1), pp. 12–12.

Mazière, C. *et al.* (2023) ‘Microbial mats as model to decipher climate change effect on microbial communities through a mesocosm study’, *Frontiers in Microbiology*, 14, p. 1039658. Available at: https://doi.org/10.3389/fmicb.2023.1039658.

McAllister, S.M., Moore, R.M. and Chan, C.S. (2018) ‘ZetaHunter, a Reproducible Taxonomic Classification Tool for Tracking the Ecology of the Zetaproteobacteria and Other Poorly Resolved Taxa’, *Microbiology Resource Announcements*, 7(7), pp. 1–3. Available at: https://doi.org/10.1128/mra.00932-18.

McMurdie, P.J. and Holmes, S. (2013) ‘phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data’, *PLoS ONE*. Edited by M. Watson, 8(4), p. e61217. Available at: https://doi.org/10.1371/journal.pone.0061217.

McNichol, J. *et al.* (2021) ‘Evaluating and Improving Small Subunit rRNA PCR Primer Coverage for Bacteria, Archaea, and Eukaryotes Using Metagenomes from Global Ocean Surveys’, *mSystems*. Edited by J.A. Gilbert, 6(3), pp. e00565-21. Available at: https://doi.org/10.1128/mSystems.00565-21.

Murali, A., Bhargava, A. and Wright, E.S. (2018) ‘IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences’, *Microbiome*, 6(1), pp. 1–14. Available at: https://doi.org/10.1186/s40168-018-0521-5.

Nearing, J.T., Comeau, A.M. and Langille, M.G.I. (2021) ‘Identifying biases and their potential solutions in human microbiome studies’, *Microbiome*, 9(1), p. 113. Available at: https://doi.org/10.1186/s40168-021-01059-0.

Oksanen, J *et al.* (2022) ‘vegan: Community Ecology Package’.

Parada, A.E., Needham, D.M. and Fuhrman, J.A. (2016) ‘Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples’, *Environmental Microbiology*, 18(5), pp. 1403–1414. Available at: https://doi.org/10.1111/1462-2920.13023.

Pascoal, F., Costa, R. and Magalhães, C. (2021) ‘The microbial rare biosphere: current concepts, methods and ecological principles’, *FEMS Microbiology Ecology*, 97(1), p. fiaa227. Available at: https://doi.org/10.1093/femsec/fiaa227.

Pollock, J. *et al.* (2018) ‘The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies’, *Applied and Environmental Microbiology*. Edited by S.-J. Liu, 84(7), pp. e02627-17. Available at: https://doi.org/10.1128/AEM.02627-17.

Quast, C. *et al.* (2013) ‘The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools’, *Nucleic Acids Research*, 41(D1), pp. 590–596. Available at: https://doi.org/10.1093/nar/gks1219.

Quince, C. *et al.* (2011) ‘Removing Noise From Pyrosequenced Amplicons’, *BMC Bioinformatics*, 12. Available at: https://doi.org/10.1186/1471-2105-12-38.

Rajeev, M. *et al.* (2020) ‘Sediment-associated bacterial community and predictive functionalities are influenced by choice of 16S ribosomal RNA hypervariable region(s): An amplicon-based diversity study’, *Genomics*, 112(6), pp. 4968–4979. Available at: https://doi.org/10.1016/j.ygeno.2020.09.006.

Ramakodi, M.P. (2021) ‘Effect of Amplicon Sequencing Depth in Environmental Microbiome Research’, *Current Microbiology*, 78(3), pp. 1026–1033. Available at: https://doi.org/10.1007/s00284-021-02345-8.

Ramírez, G.A. *et al.* (2021) 'Environmental factors shaping bacterial, archaeal and fungal community structure in hydrothermal sediments of Guaymas Basin, Gulf of California', *PLOS ONE*. Edited by W.J. Brazelton, 16(9), p. e0256321. Available at: https://doi.org/10.1371/journal.pone.0256321.

Reese, A.T. and Dunn, R.R. (2018) 'Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance', *mBio*. Edited by M.J. McFall-Ngai, 9(4), pp. e01294-18. Available at: https://doi.org/10.1128/mBio.01294-18.

Resing, J.A. *et al.* (2015) 'Basin-scale transport of hydrothermal dissolved metals across the South Pacific Ocean', *Nature*, 523(7559), pp. 200–203. Available at: https://doi.org/10.1038/nature14577.

Scott, J.J., Glazer, B.T. and Emerson, D. (2017) 'Bringing microbial diversity into focus: high-resolution analysis of iron mats from the Lō'ihi Seamount', *Environmental Microbiology*, 19(1), pp. 301–316. Available at: https://doi.org/10.1111/1462-2920.13607.

Shakya, M. *et al.* (2013) 'Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities: Metagenomic and rRNA diversity characterization', *Environmental Microbiology*, 15(6), pp. 1882–1899. Available at: https://doi.org/10.1111/1462-2920.12086.

Soriano-Lerma, A. *et al.* (2020) 'Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples', *Scientific Reports*, 10(1), p. 13637. Available at: https://doi.org/10.1038/s41598-020-70141-8.

Speth, D.R. *et al.* (2022) 'Microbial communities of Auka hydrothermal sediments shed light on vent biogeography and the evolutionary history of thermophily', *The ISME Journal*, 16(7), pp. 1750–1764. Available at: https://doi.org/10.1038/s41396-022-01222-x.

Stromecki, A. *et al.* (2022) 'Unexpected diversity found within benthic microbial mats at hydrothermal springs in Crater Lake, Oregon', *Frontiers in Microbiology*, 13, p. 876044. Available at: https://doi.org/10.3389/fmicb.2022.876044.

Vander Roost, J., Thorseth, I.H. and Dahle, H. (2017) 'Microbial analysis of Zetaproteobacteria and co-colonizers of iron mats in the Troll Wall Vent Field, Arctic Mid-Ocean Ridge', *PLOS ONE*. Edited by T.-Y. Chiang, 12(9), p. e0185008. Available at: https://doi.org/10.1371/journal.pone.0185008.

Walters, W. *et al.* (2015) 'Transcribed Spacer Marker Gene Primers for Microbial Community Surveys', *mSystems*, 1(1), pp. e0009-15. Available at: https://doi.org/10.1128/mSystems.00009-15.Walters.

Wang, H. *et al.* (2021) 'The characteristics of Fe speciation and Fe-binding ligands in the Mariana back-arc hydrothermal plumes', *Geochimica et Cosmochimica Acta*, 292, pp. 24–36. Available at: https://doi.org/10.1016/j.gca.2020.09.016.

Wang, W. *et al.* (2021) 'Behavior of iron isotopes in hydrothermal systems: Beebe and Von Damm vent fields on the Mid-Cayman ultraslow-spreading ridge', *Earth and Planetary Science Letters*, 575, p. 117200. Available at: https://doi.org/10.1016/j.epsl.2021.117200.

Wang, Y. and Qian, P.Y. (2009) 'Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies', *PLoS ONE*, 4(10). Available at: https://doi.org/10.1371/journal.pone.0007401.

Wear, E.K. *et al.* (2018) 'Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton', *Environmental Microbiology*, 20(8), pp. 2709–2726. Available at: https://doi.org/10.1111/1462-2920.14091.

Wheat, C.G. *et al.* (2000) 'Continuous sampling of hydrothermal fluids from Loihi Seamount after the 1996 event', *Journal of Geophysical Research: Solid Earth*, 105(B8), pp. 19353–19367. Available at: https://doi.org/10.1029/2000jb900088.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer (Use R!). Available at: https://doi.org/10.1007/978-3-319-24277-4.

Wickham, H. *et al.* (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. Available at: https://doi.org/10.21105/joss.01686.

Yilmaz, P. *et al.* (2014) 'The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks', *Nucleic Acids Research*, 42(D1), pp. D643–D648. Available at: https://doi.org/10.1093/nar/gkt1209.

Table 1. Sample metadata.

Collection ID	Sample Name	Primer	Location	Latitude	Longitude	Depth (m)
J2-674-Green	Bact674Green	V3V4	Pohaku	18.90134	-155.2582	1179
	Uni674Green	V4V5				
J2-674-Blue	Bact674Blue	V3V4	Pohaku	18.90155	-155.2582	1182
	Uni674Blue	V4V5				
J2-677-Black	Bact677Black	V3V4	Hiolo South	18.90554	-155.257	1270
	Uni677Black	V4V5				
J2-675-Black	Bact675Black	V3V4	Hiolo South	18.9056	-155.257	1272
	Uni675Black	V4V5				
J2-677-Green	Bact677Green	V3V4	Hiolo North	18.90645	-155.257	1300
	Uni677Green	V4V5				
J2-677-Blue	Bact677Blue	V3V4	Hiolo North	18.90645	-155.2569	1300
	Uni677Blue	V4V5				
J2-676-Black	Bact676Black	V3V4	Lohiau	18.90886	-155.2577	1175
	Uni676Black	V4V5				
J2-676-Green	Bact676Green	V3V4	Lohiau	18.90887	-155.2577	1175
	Uni676Green	V4V5				

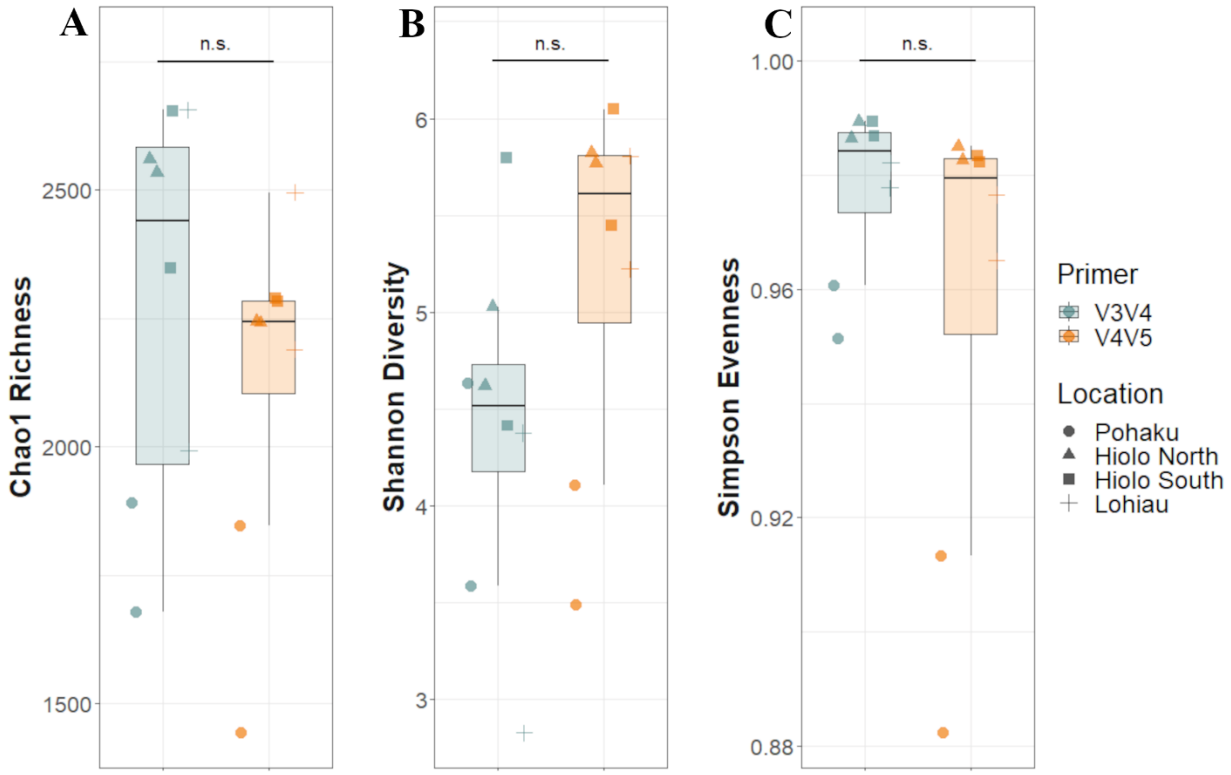


Figure 1. Alpha diversity per primer set. (A) Chao1 Richness, (B) Shannon Diversity, and (C) Simpson Evenness estimates for each microbial mat community. Different primer sets represented by different colors; different locations represented by different shapes. NB: the differences in y-axis values are reflected for each diversity index. Significance was assessed using analysis of variance for normally distributed data (Shannon and Chao1) or Kruskal-Wallis for non-normally distributed data (Simpson).

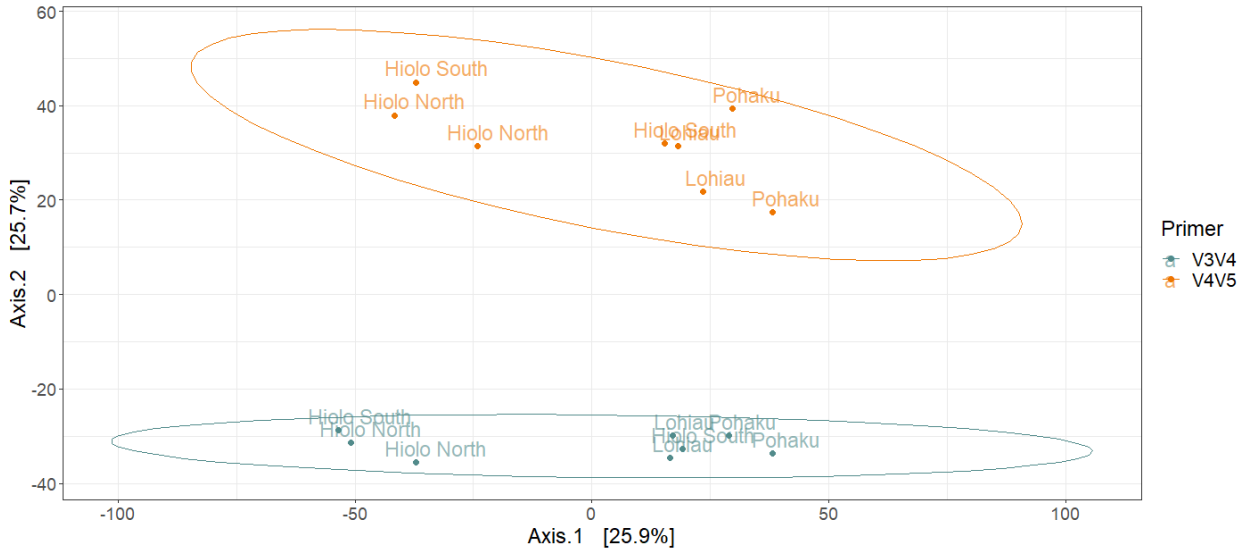


Figure 2. Principal coordinate analysis of microbial mat communities by primer set, with ellipses indicative of 95% confidence interval.

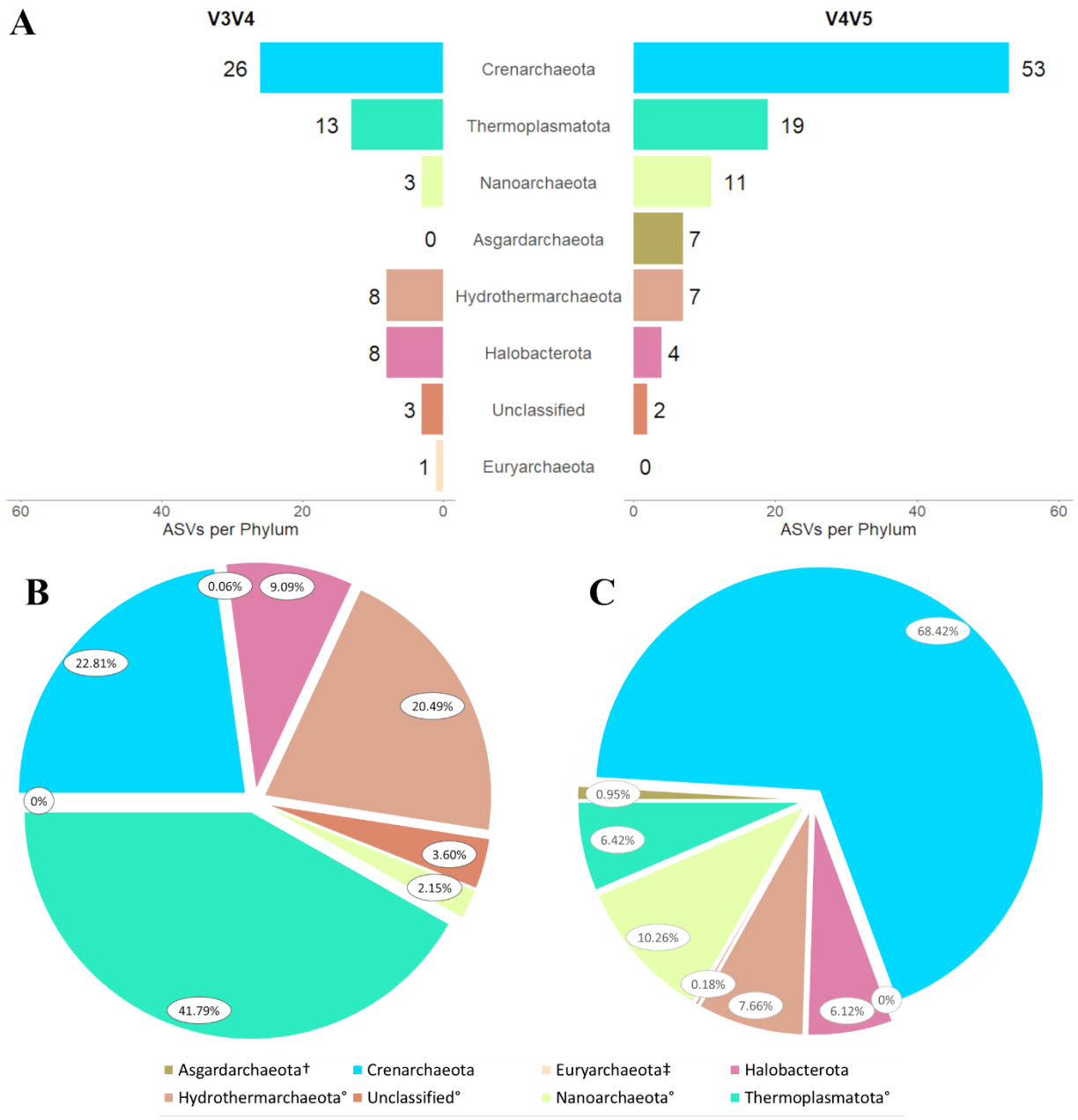


Figure 3. (A) Diversity in domain Archaea captured by each primer set as described via ASVs per phylum and as described via phylum relative abundance for (B) primer set V3V4 and (C) primer set V4V5. In (B) and (C) legend, † denotes taxa present in V4V5 but absent in V3V4, ‡ denotes taxa present in V3V4 but absent in V4V5, ° denotes taxa that differ in relative abundance by at least one order of magnitude between primer sets. NB: in (B), the non-visible 0.06% slice represents the phylum Euryarchaeota.

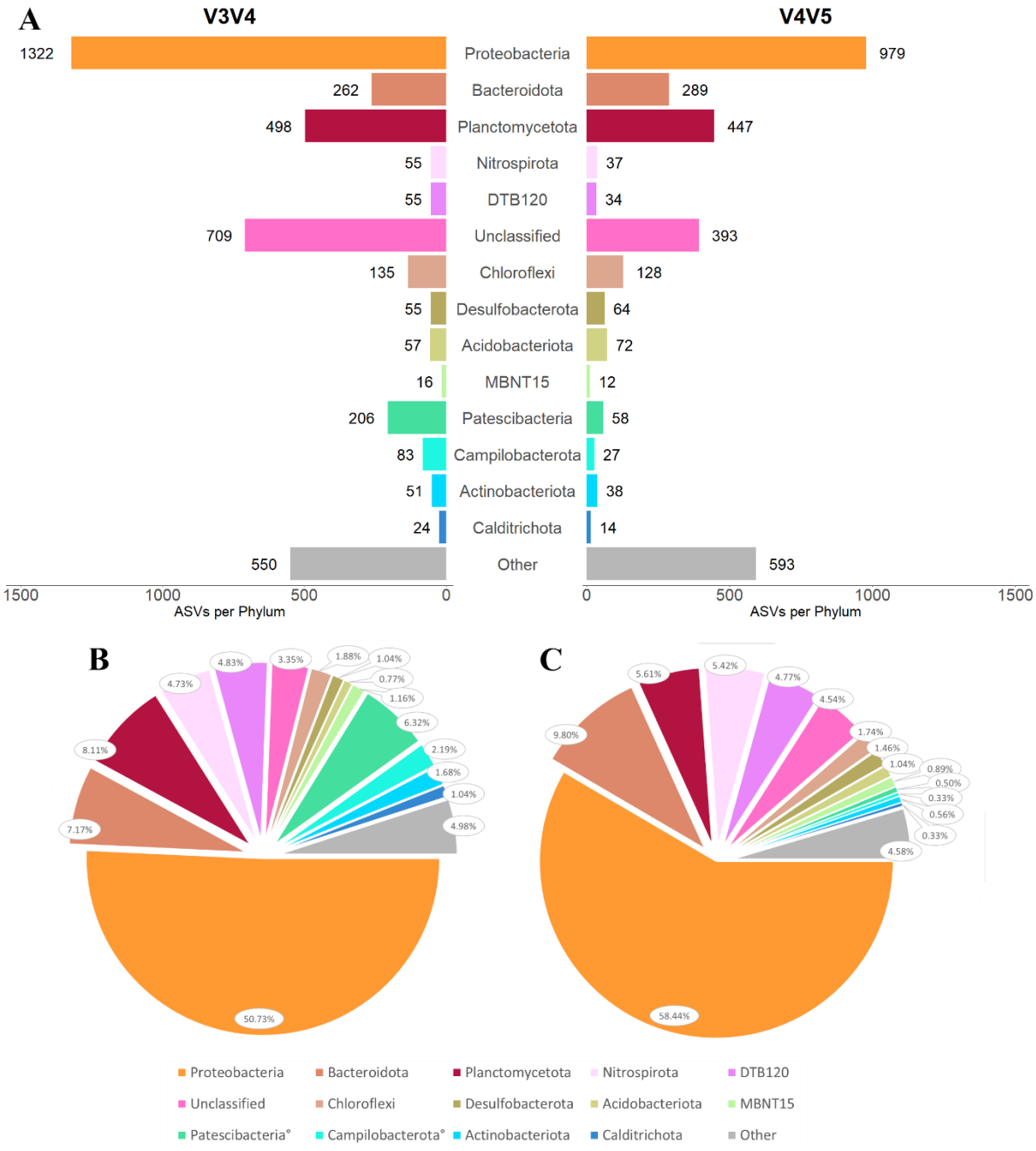


Figure 4. (A) Diversity in domain Bacteria captured by each primer set as described via ASVs per phylum for phyla representing greater than 1% relative abundance in either primer set and via phylum relative abundance for (B) primer set V3V4 and (C) primer set V4V5. In (B) and (C) legend, ° denotes taxa that differ in relative abundance by at least one order of magnitude between primers.

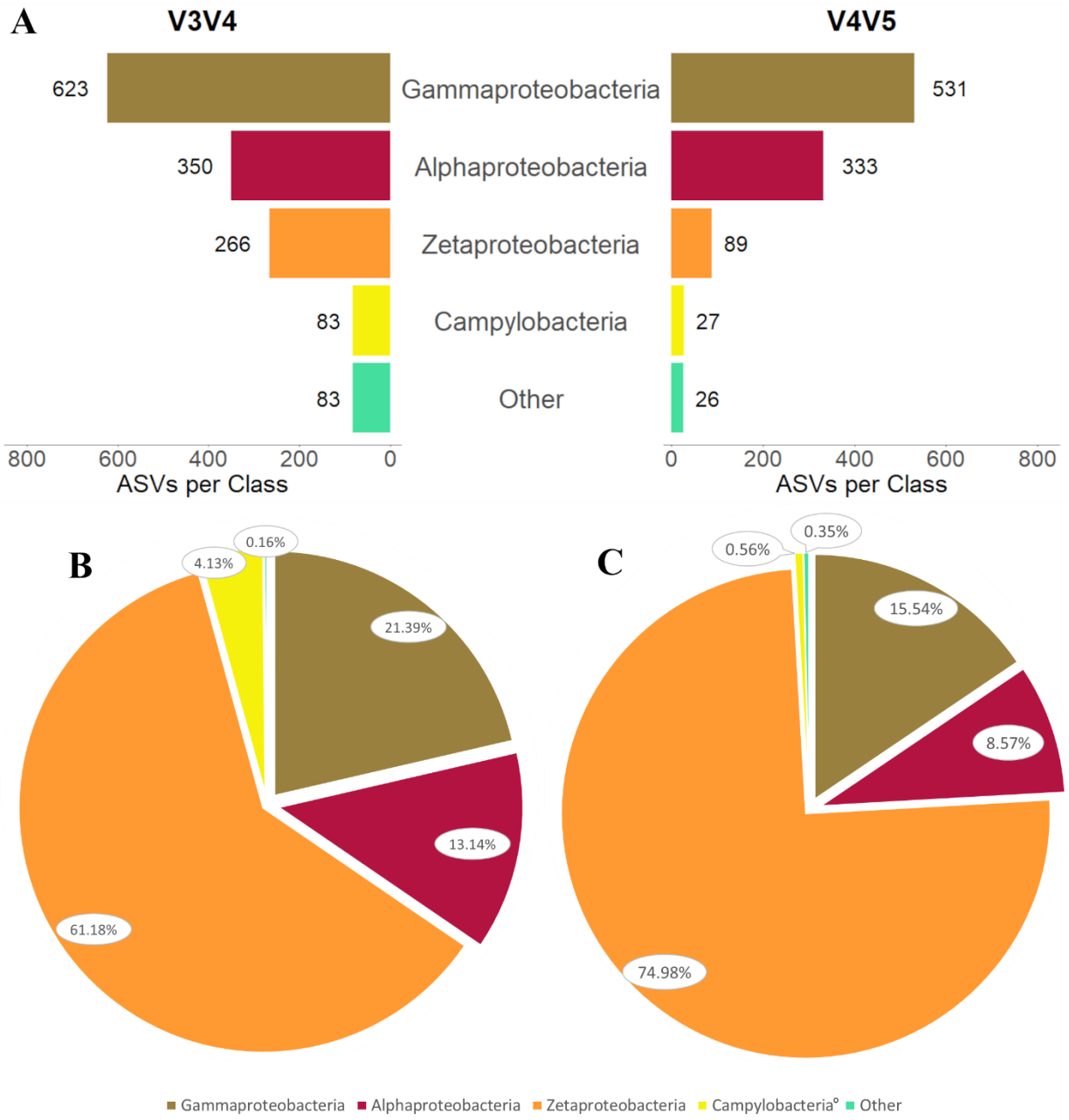


Figure 5. (A) Diversity in phylum Proteobacteria captured by each primer set as described via ASVs per class and via class relative abundance for (B) primer set V3V4 and (C) primer set V4V5. In (B) and (C) legend, ° denotes taxa that differ in relative abundance by at least one order of magnitude between primers. NB: Class Campylobacteria in Phylum Campylobacterota is included in analysis of phylum Proteobacteria here since it was formerly class Epsilonproteobacteria.

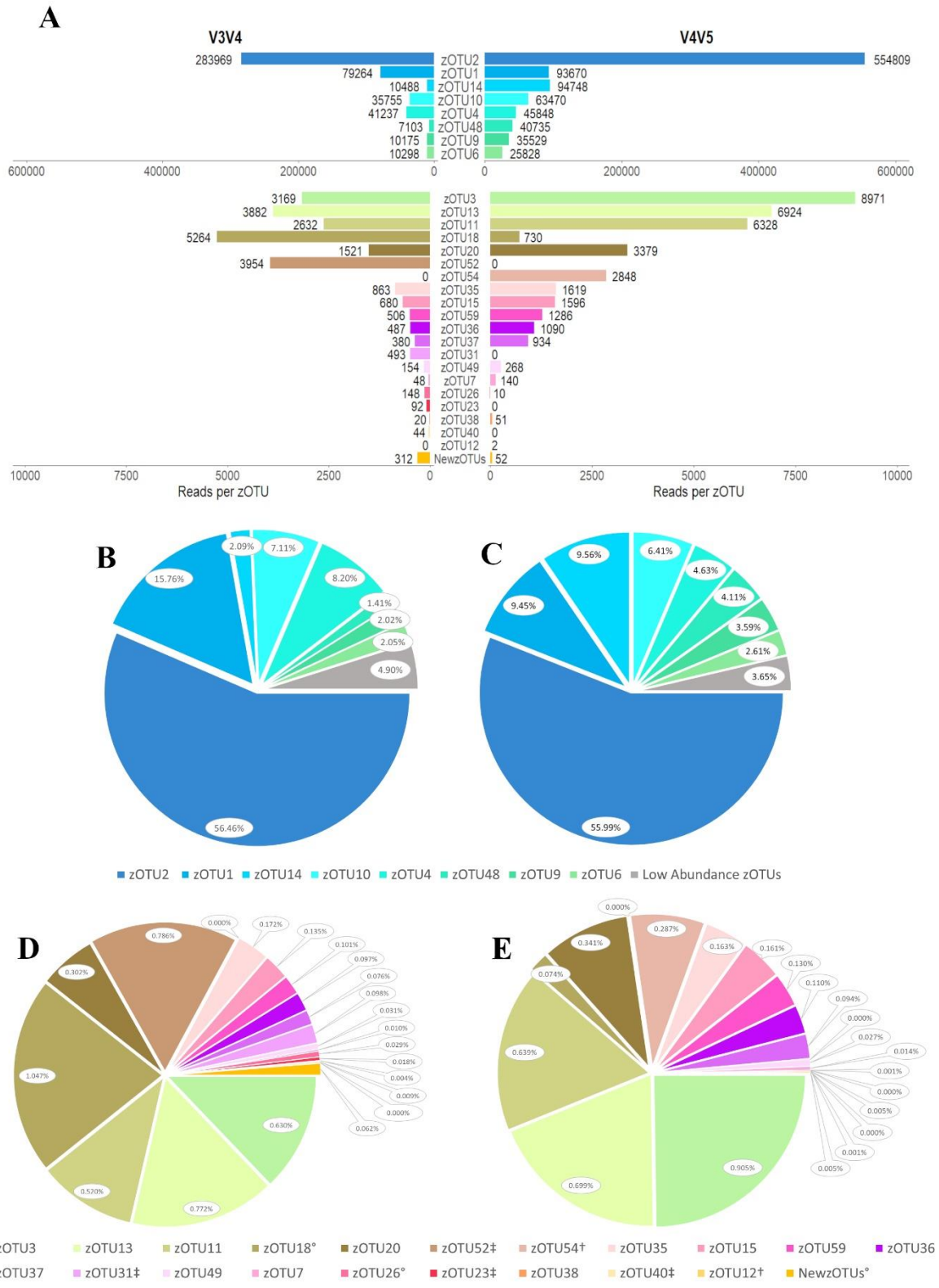

A
V3V4
V4V5
283969 zOTU2 554809
79264 zOTU1 93670
10488 zOTU14 94748
35755 zOTU10 63470
41237 zOTU4 45848
7103 zOTU48 40735
10175 zOTU9 35529
10298 zOTU6 25828
600000 400000 200000 0 0 200000 400000 600000
3169 zOTU3 8971
3882 zOTU13 6924
2632 zOTU11 6328
5264 zOTU18 730
1521 zOTU20 3379
3954 zOTU52 0
0 zOTU54 2848
863 zOTU35 1619
680 zOTU15 1596
506 zOTU59 1286
487 zOTU36 1090
380 zOTU37 934
493 zOTU31 0
154 zOTU49 268
48 zOTU7 140
148 zOTU26 10
92 zOTU23 0
20 zOTU38 51
44 zOTU40 0
0 zOTU12 2
312 NewzOTUs 52
10000 7500 5000 2500 0 0 2500 5000 7500 10000
Reads per zOTU
Reads per zOTU
B
2.09%
7.11%
15.76%
8.20%
1.41%
2.02%
2.05%
4.90%
56.46%
C
9.56%
6.41%
4.63%
9.45%
4.11%
3.59%
2.61%
3.65%
55.99%
zOTU2 zOTU1 zOTU14 zOTU10 zOTU4 zOTU48 zOTU9 zOTU6 Low Abundance zOTUs
D
0.000%
0.172%
0.135%
0.786%
0.101%
0.097%
0.302%
0.076%
0.098%
0.031%
0.010%
0.029%
0.018%
1.047%
0.004%
0.009%
0.000%
0.062%
0.630%
0.520%
0.772%
E
0.000%
0.287%
0.341%
0.163%
0.161%
0.130%
0.074%
0.110%
0.094%
0.000%
0.027%
0.014%
0.639%
0.001%
0.000%
0.005%
0.000%
0.001%
0.005%
0.905%
0.699%
zOTU3 zOTU13 zOTU11 zOTU18° zOTU20 zOTU52‡ zOTU54† zOTU35 zOTU15 zOTU59 zOTU36
zOTU37 zOTU31‡ zOTU49 zOTU7 zOTU26° zOTU23‡ zOTU38 zOTU40‡ zOTU12† NewzOTUs°

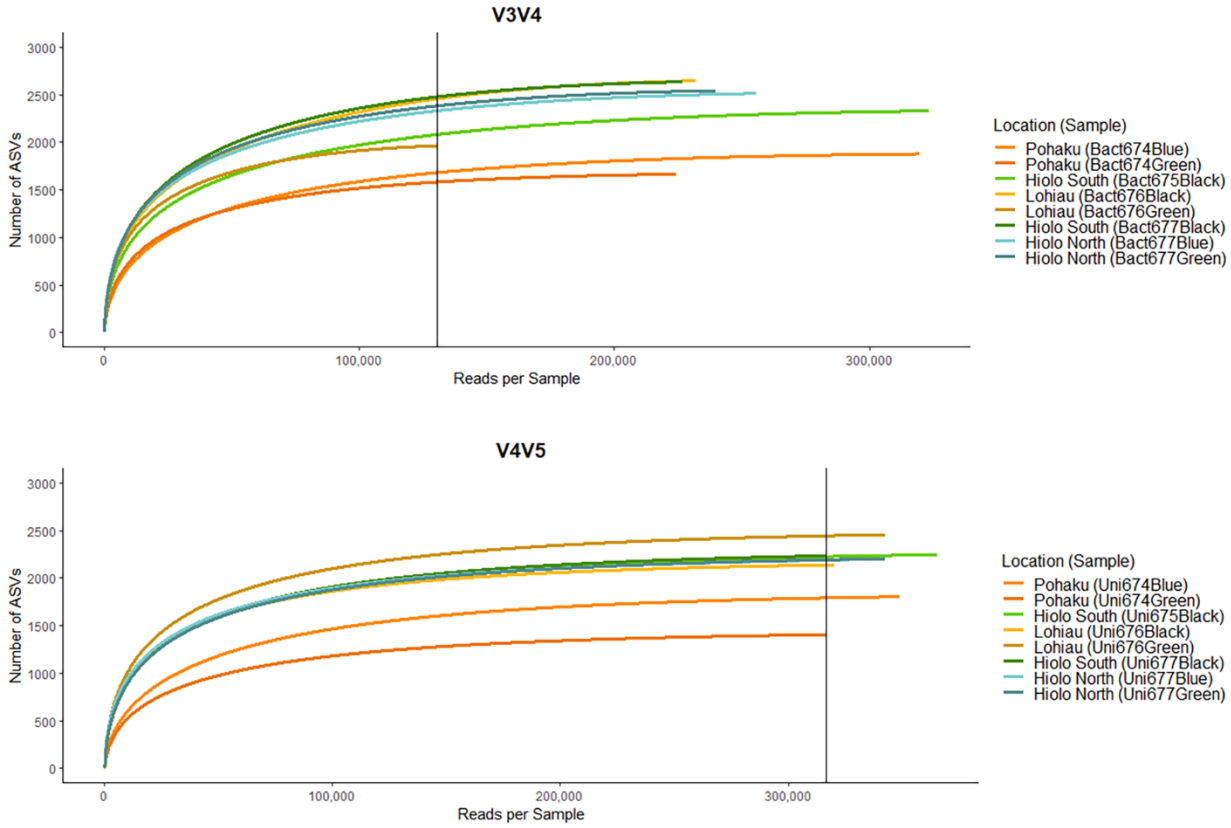
Figure 6. (A) Diversity in class Zetaproteobacteria captured by each primer set as described via reads per zOTU designation, NB: the break in the y-axis represents a change in scale with respect to reads per zOTU. Zetaproteobacteria diversity is also described per primer set via relative abundance of both high-abundance, (i.e., greater than 10^4 reads per taxon for B and C), and low-abundance, (i.e., fewer than 10^4 reads per taxon for D and E). In (D) and (E) legend, † denotes taxa present in V4V5 but absent in V3V4, ‡ denotes taxa present in V3V4 but absent in V4V5, and ° denotes taxa that differed in relative abundance by at least one order of magnitude between primer sets.

Supplemental Table 1. Read tracking through the DADA2 pipeline for the V3V4 samples.

V3V4 Sample	input	filtered	denoised	merged	nonchim
Bact674Green (Pohaku)	536524	312116 41.83% reads lost from previous step 41.83% reads lost total	305827 2.01% reads lost from previous step 43.00% reads lost total	287901 5.86% reads lost from previous step 46.34% reads lost total	224147 22.14% reads lost from previous step **58.22% reads lost total**
Bact674Blue (Pohaku)	704150	425615 39.56% reads lost from previous step 39.56% reads lost total	418378 1.70% reads lost from previous step 40.58% reads lost total	396312 5.27% reads lost from previous step 43.72% reads lost total	319630 19.35% reads lost from previous step **54.61% reads lost total**
Bact677Black (Hiolo South)	671908	352141 47.59% reads lost from previous step 47.59% reads lost total	332288 5.64% reads lost from previous step 50.55% reads lost total	289586 12.85% reads lost from previous step 56.90% reads lost total	226794 21.68% reads lost from previous step **66.25% reads lost total**
Bact675Black (Hiolo South)	764345	430823 43.64% reads lost from previous step 43.64% reads lost total	415302 3.60% reads lost from previous step 45.67% reads lost total	381774 8.07% reads lost from previous step 50.05% reads lost total	323532 15.26% reads lost from previous step **57.67% reads lost total**
Bact677Green (Hiolo North)	710932	343559 51.67% reads lost from previous step 51.67% reads lost total	326062 5.09% reads lost from previous step 54.14% reads lost total	290176 11.01% reads lost from previous step 59.18% reads lost total	239881 17.33% reads lost from previous step **66.26% reads lost total**
Bact677Blue (Hiolo North)	678461	346545 48.92% reads lost from previous step 48.92% reads lost total	327689 5.44% reads lost from previous step 51.70% reads lost total	295553 9.81% reads lost from previous step 56.44% reads lost total	255559 13.53% reads lost from previous step **62.33% reads lost total**
Bact676Black (Lohiau)	697317	381310 45.32% reads lost from previous step 45.32% reads lost total	358161 6.07% reads lost from previous step 48.64% reads lost total	315104 12.02% reads lost from previous step 54.81% reads lost total	232219 26.30% reads lost from previous step **66.70% reads lost total**
Bact676Green (Lohiau)	523552	237707 54.60% reads lost from previous step 54.60% reads lost total	217428 8.53% reads lost from previous step 58.47% reads lost total	183625 15.55% reads lost from previous step 64.93% reads lost total	130311 29.03% reads lost from previous step **75.11% reads lost total**

Supplemental Table 2. Read tracking through the DADA2 pipeline for the V4V5 samples.

V4V5 Sample	input	filtered	denoised	merged	nonchim
Uni674Green (Pohaku)	631884	354032 43.97% reads lost from previous step 43.97% reads lost total	351996 0.58% reads lost from previous step 44.29% reads lost total	345154 1.94% reads lost from previous step 45.38% reads lost total	316986 8.16% reads lost from previous step **49.83% reads lost total**
Uni674Blue (Pohaku)	676630	390580 42.28% reads lost from previous step 42.28% reads lost total	387773 0.72% reads lost from previous step 42.69% reads lost total	378857 2.30% reads lost from previous step 44.01% reads lost total	348909 7.91% reads lost from previous step **48.43% reads lost total**
Uni677Black (Hiolo South)	731706	400832 45.22% reads lost from previous step 45.22% reads lost total	390065 2.69% reads lost from previous step 46.69% reads lost total	360354 7.62% reads lost from previous step 50.75% reads lost total	316860 12.07% reads lost from previous step **56.70% reads lost total**
Uni675Black (Hiolo South)	713753	387859 45.66% reads lost from previous step 45.66% reads lost total	384187 0.95% reads lost from previous step 46.17% reads lost total	374436 2.54% reads lost from previous step 47.54% reads lost total	365477 2.39% reads lost from previous step **48.80% reads lost total**
Uni677Green (Hiolo North)	739724	395620 46.52% reads lost from previous step 46.52% reads lost total	388541 1.80% reads lost from previous step 47.47% reads lost total	369874 4.80% reads lost from previous step 50.0% reads lost total	342737 7.34% reads lost from previous step **53.67% reads lost total**
Uni677Blue (Hiolo North)	765667	390372 49.02% reads lost from previous step 49.02% reads lost total	379535 2.78% reads lost from previous step 50.43% reads lost total	354440 6.61% reads lost from previous step 53.71% reads lost total	327206 7.68% reads lost from previous step **57.27% reads lost total**
Uni676Black (Lohiau)	684834	359354 47.53% reads lost from previous step 47.53% reads lost total	354785 1.27% reads lost from previous step 48.19% reads lost total	339673 4.26% reads lost from previous step 50.4% reads lost total	320145 4.99% reads lost from previous step **53.25% reads lost total**
Uni676Green (Lohiau)	711945	400198 43.79% reads lost from previous step 43.79% reads lost total	392078 2.03% reads lost from previous step 44.93% reads lost total	372252 5.06% reads lost from previous step 47.71% reads lost total	342638 7.96% reads lost from previous step **51.87% reads lost total**



Supplemental Figure 1. Rarefaction curves for each microbial mat community, colored by sample location, for each primer set. Vertical lines denote the minimum read depth.