Depth drives the distribution of microbial ecological functions in the coastal western Antarctic Peninsula

Avishek Dutta[1,2,3*], Elizabeth Connors[1], Rebecca Trinh[4], Natalia Erazo[1], Srishti Dasarathy[1], Hugh W. Ducklow[4], Deborah K. Steinberg[5], Oscar M. Schofield[6], Jeff S. Bowman[1,7]

[1]Integrative Oceanography Division, Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92093, USA

[2]Department of Geology, University of Georgia, Athens, GA 30602, USA

[3]Savannah River Ecology Laboratory, University of Georgia, Aiken, SC 29808, USA

[4]Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA

[5]Department of Biological Science, Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA 23062, USA

[6]Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA

[7]Center for Microbiome Innovation, UC San Diego, La Jolla, CA 92093, USA

Corresponding Author:

Avishek Dutta

E-mail ID: avishek.dutta@uga.edu

Abstract

The Antarctic marine environment is a dynamic ecosystem where microorganisms play an important role in key biogeochemical cycles. Despite the role that microbes play in this ecosystem, little is known about the genetic and metabolic diversity of Antarctic marine microbes. In this study we leveraged DNA samples collected by the Palmer Long Term Ecological Research (LTER) project to sequence shotgun metagenomes from 48 key locations across the marine ecosystem of the western Antarctic Peninsula. We developed an *in silico* metagenomics pipeline (iMAGine) for processing metagenomic data and constructing metagenome-assembled genomes (MAGs), identifying a diverse genomic repertoire related to the carbon, sulfur, and nitrogen cycles. A novel analytical approach based on gene coverage was used to understand the differences in microbial community functions across depth and region. Our results showed that microbial community functions were partitioned based on depth. Bacterial members harbored diverse genes for carbohydrate transformation, indicating the availability of processes to convert complex

carbons into simpler bioavailable forms. We generated 137 dereplicated MAGs giving us a new perspective on the role of prokaryotes in the coastal western Antarctic Peninsula. In particular, the presence of otrophic prokaryotes capable of autotrophic and heterotrophic lifestyles indicated a metabolically flexible community, which we hypothesize enables survival under rapidly changing conditions. Overall, the study identified key microbial community functions and created a valuable sequence library collection for future Antarctic genomics research.

1. Introduction

Marine microorganisms play an important role in regulating biogeochemical cycles (Green et al., 2008). They are key drivers of the transformation of carbon-, nitrogen-, and sulfur-containing compounds in the environment. Changes in environmental conditions impact microbial communities, which in turn exert control over many environmental parameters (Thompson et al., 2017; Dutta et al., 2022). The ecological outcomes of the rapid environmental change are well documented for the western Antarctic Peninsula (wAP) (Meredith and King, 2005; Clarke et al., 2007; Bowman et al., 2016, 2017), home to multiple long-term observing programs. Large shifts in bacterial production relative to primary production signal radically different outcomes for primary production from one year to another (Bowman et al., 2016). The metabolic potential of the heterotrophic bacterial community is presumed to play a strong role in determining what primary production gets recycled by the microbial food web. However, we know little about the genomic makeup of the bacteria and archaea responsible for bacterial production and other marine microbial processes along the wAP (Bowman et al., 2018).

Heterotrophic bacterial populations are intimately linked to phytoplankton blooms and play an essential role in the transformation of phytoplankton-derived organic matter (Buchan et al., 2014). Phytoplankton are a direct source of dissolved organic carbon (DOC) for heterotrophic bacteria in the photic zone. Below the photic zone, heterotrophs reprocess DOC and degrade sinking particles to generate new DOC. The timing of the seasonal phytoplankton bloom, its composition, and its intensity are strongly influenced by physical processes along the wAP. For example, conditions that favor large diatoms are thought to transfer carbon more efficiently to krill and upper trophic-level consumers (Saba et al., 2014). Alternatively, strong winds and reduced sea ice cover can lead to lower levels of primary production and smaller phytoplankton cells, in turn leading to high rates of bacterial production compared to primary production (Bowman et al., 2016). Recent trends and

future climate scenarios suggest an increase in wind and a reduction in sea ice for the Antarctic peninsula (Siegert et al., 2019) and, presumably, a strengthened microbial food web. To better understand the metabolic capabilities of wAP marine bacterial communities, we applied metagenomics to a historic sample library of microbial DNA collected by the Palmer Long Term Ecological Research (LTER) project to better understand how bacterial communities will respond to future environmental change along the wAP.

Most primary production along the wAP is attributed to eukaryotic phytoplankton (Schofield et al., 2018; Lin et al., 2021). However, dark carbon fixation is likely to be a significant process below the photic zone and during the polar night. Though well appreciated for the global ocean (Baltar and Herndl, 2019), surprisingly little is known about the distribution of prokaryotic carbon fixation mechanisms in the Antarctic marine environment. Previous analysis of fosmid libraries from contrasting summer and winter communities along the wAP identified gammaproteobacterial sulfur-oxidizing (GSO) chemolithotrophs (Grzymski et al., 2012). Other works using 16S rRNA gene surveys have shown these taxa to be widely distributed in the coastal Antarctic (Bowman and Deming, 2017; Bowman et al., 2017). Alternate prokaryotic carbon fixation strategies for the wAP may rely on energy obtained from nitrification (Bowman et al., 2016). This study aimed to understand the microbial community functions in the marine ecosystem of coastal wAP. We applied a novel analytical approach based on the gene coverage to investigate the distribution of genes and reconstructed metagenome-assembled genomes (MAGs) to understand the pathways associated with prokaryotic carbon fixation and utilization. We combined observations of genes diagnostic of carbon fixation with genes for catabolic processes to identify autotrophic, mixotrophic, and heterotrophic guilds among wAP marine prokaryotes.

2. Methods

2.1.Sample collection

Forty-eight samples were selected covering different depth horizons (0-275 m) on the Palmer LTER sampling grid (lines 100-600, covering latitudes 67.566533 °S to 63.96565 °S) (Waters and Smith, 1992) and locations of special significance such as Palmer Canyon, Armstrong Reef, and the coastal LTER time-series near Palmer Station (Figure 1). The selection of samples was made based on three different parameters: (i) higher abundances of unclassified taxa based on 16S rRNA amplicon sequence data, (ii) depth profile, and (iii) variations in latitude. A detailed description of

the samples used in this study is given in Supplementary Table S1. Forty-two samples were collected during the austral summer of 2019-2020, and four samples were collected during the austral summer of 2018-2019. The remaining two samples were collected in January 2017 and November 2018. We grouped the sampled into three depth horizons: shallow (0-40 m), medium (50-75 m), and deep (100-275 m). The depth categorization was done based on mixed layer depth, temperature, and salinity as reported previously in Schofield et al., 2018 and Seyitmuhammedov et al., 2022. There were 23 samples collected from the shallower depth, whereas 11 and 14 samples were collected from the medium-depth and deeper horizons, respectively. The samples were also categorized into three regions based on Palmer LTER lines: Northern (Line 400 and north of Line 400), Southern (south of Line 400), and Palmer Canyon (near Palmer Station on Anvers Island) (Supplementary Table S1). There were 14 samples collected from the Northern region, whereas 16 and 18 samples were collected from Palmer Canyon and the Southern region, respectively. For each sample, approximately 1 L of seawater was filtered through a sterile 0.2 µm Supor membrane disc filter (Pall Corporation, Port Washington, NY, USA) and stored at −80 ºC until extraction.

2.2.DNA extraction, sequencing, and metagenome analysis

DNA was extracted from the filters using the MagMAX Microbiome Ultra nucleic acid extraction kit and KingFisher Flex extraction system following the manufacturer's protocols. The extracted DNA was sequenced at the UC San Diego Microbiome Core for shotgun metagenome sequencing on the Illumina NovaSeq platform. Sequencing was done across multiple lanes in two runs (24 samples for each run). The average depth of sequencing for the second run (an average of ~427 million paired-end reads per sample) was higher compared to the first run (an average of ~55 million paired-end reads per sample) to facilitate a separate analysis that will be reported elsewhere. To avoid memory limitation, samples from the second run were down-sampled using reformat.sh, a script from the BBMap package (Bushnell, 2014), to ~55 million paired-end reads per sample. The raw metagenomic data for 48 samples were processed and analyzed using the *in silico* Metagenomics Pipeline (iMAGine) using default parameters (https://github.com/avishekdutta14/iMAGine). iMAGine uses fastp (Chen et al., 2018) for filtering, metaSPAdes (Nurk et al., 2017) for assembling the reads, QUAST (Gurevich et al., 2013) for analyzing the assembly quality, bwa-mem (v0.7.17) for aligning the raw reads to the assembly (Li, 2013), samtools for modifying alignment files (Li et al., 2009), metabat2 (Kang et al., 2019)

for binning contigs, and checkM (Parks et al., 2015) for quality assessment of the bins. The assembled contigs from iMAGine were used for further analyses.

Genes were predicted from contigs with Prodigal v2.6.3 (Hyatt et al., 2010) using the 'meta' flag. The predicted genes were annotated using emapper v2.1.5 (Cantalapiedra et al., 2021) based on a Diamond search (Buchfink et al., 2014). The following arguments were used for search filtering in emapper: --evalue 0.001 (e-value threshold), --score 60 (minimum hit bit score), --pident 40 (minimum percentage identity), --query_cover 20 (minimum percentage query coverage), and --subject_cover 20 (minimum percentage subject coverage). The database used for annotation was eggNOG DB v5.0.2 (Huerta-Cepas et al., 2019). this study, taxfin.sh (a part of iMAGine) was used for keeping genes affiliated to domain bacteria and archaea. Coverage of each gene (average gene fold) was determined using gene_fold_counter.sh (a part of iMAGine), which takes in the alignment map file (sam output from iMAGine), removes unmapped reads and reads mapped to multiple locations using samtools (with parameters -F 0x904), uses pileup.sh script from the BBMap package to calculate the average coverage of the contigs and maps back average contig coverage to the genes on those contigs. To enable comparison across metagenomes, all the genes were scaled based on Eqt 1:

$$Normalized\ gene\ coverage\ of\ a\ sample = \frac{Total\ average\ gene\ coverage}{Total\ \ rpoB\ (K03043)\ coverage} \quad (1)$$

For gene-specific analysis, KEGG orthologs from the emapper outputs were considered. Genes mapping to more than one ortholog were not considered for the analysis. Key genes involved in different processes of carbon, nitrogen, and sulfur cycles were selected based on KEGG pathways, whereas genes involved in carbohydrate transformation were selected based on previously published literature (Bergauer et al., 2018). Detailed gene inventories, along with the associated pathways considered in this study, are mentioned in Supplementary Figures S1, S2, S3, and S4. The KEGG ortholog for *rpoB* i.e. K03043 was used as a reference for normalization in this study. The genes involved in methane oxidation considered in this study (*pmoA-amoA*, *pmoB-amoB*, and *pmoC-amoC*) also play an important role in ammonia oxidation (a key step in nitrification). Similarly, it has been seen before that the variations of the same gene responsible for dissimilatory sulfate reduction (considered in this study) are also involved in sulfur oxidation (Loy et al., 2009).

Bins with completeness higher than 70% and contamination lower than 5% were considered MAGs. Similar cutoffs were used in a previous study (Parks et al., 2017). For further analysis, MAGs were dereplicated with dRep v3.2.2 (Olm et al., 2017) using a secondary ANI threshold of 0.96 and goANI as the algorithm for secondary clustering. The minimum genome completeness was set to 70%, and the maximum genome contamination was set to 5% for dereplication. All the other dereplication parameters were kept default. The taxonomy of the dereplicated MAGs was assessed using the GTDB-Tk v1.5.0 based on reference database version release 202 (Chaumeil et al., 2020). The multiple sequence alignment output of domain-specific marker genes from GTDB-Tk was used for constructing a phylogenomic tree with RAxML-ng (Kozlov et al., 2019). Tree construction used the WAG amino-acid substitution model and MAGs belonging to Chloroflexota and Thermoproteota as outgroups for bacteria- and archaea-specific phylogenomic trees, respectively. The default number of starting trees was used (ten random and ten parsimony-based) and one thousand bootstrap replicates were used for confidence scoring of the final tree.

The number of reads mapped to each MAG was calculated using the mag_abund.py script present in the iMAGine repository (https://github.com/avishekdutta14/iMAGine/blob/main/Utilities/mag_abund.py), and the average reads per secondary cluster (as obtained from dRep analyses) was obtained to understand average reads mapped per dereplicated MAGs.

2.3.Determination of putative metabolic lifestyle in MAGs

Functional annotation of MAGs was conducted using two different methods. To identify functional guilds, genes were predicted using Prodigal v2.6.3 and annotated using Ghost Koala (Kanehisa et al., 2016), whereas DRAM 1.4.0 was used to assess pathway completeness and categorize different microbial metabolisms (Shaffer et al., 2020). Though it is hard to ascertain the exact metabolic lifestyle of the MAGs given the limitation imposed by genome completeness, we described the putative metabolisms based on binations of diagnostic metabolic pathways. All the annotations for metabolic lifestyle assessment were based on DRAM analysis. Glucose utilization (GU) and carbon fixation (CF) pathways having greater than 70% completeness in a MAG were considered in this study. GU, indicative of the use of exogenous fixed carbon, was assessed by the presence of the Embden Meyerhof pathway and/or the Entner-Doudoroff pathway. CF was assessed based on the presence of either of the three carbon fixing pathways (3-Hydroxypropionate

bicycle, Arnon–Buchanan cycle, and Calvin cycle) found in the MAGs. Sulfur oxidation (SO) capability was determined based on the presence of the SOX system and/or the presence of the *dsr* gene. Nitrification (NI) capability was determined based on the presence of either ammonia oxidation genomic repertoire and/or gene involved in the conversion of nitrite to nitrate. Denitrification (DNR) capability was determined based on the presence of either one of the genes involved in the following processes: (i) genes involved in the conversion of nitrate to nitrite, (ii) genes involved in the conversion of nitrite to nitric oxide, and/or (iii) genes involved in the conversion of nitric oxide to nitrous oxide. We described each MAG as a putative autotroph, mixotroph, or heterotroph based on the presence of diagnostic metabolic pathways. Autotrophs were determined by the presence of CF and absence of GU, heterotrophs by the presence of GU and absence of CF, and mixotrophs by the presence of both GU and CF.

2.4.Statistical analyses

All the statistical analyses were carried out in R and R Studio (Team, 2015). Principal Coordinates Analysis (PCoA) of Bray-Curtis dissimilatory based on the normalized gene coverage across different samples was performed using the phyloseq package (McMurdie and Holmes, 2013). For this PCoA, we considered key genes responsible for carbon, nitrogen, sulfur, and carbohydrate transformations (Supplementary Figure S1, S2, S3, S4). A PERMANOVA test was conducted using the 'adonis2' function of the Vegan package (Oksanen et al., 2007) to test whether different sample groups had different centroids. The average distance from the median was calculated based on a dispersion test using the 'betadisper' function of the Vegan package, which was followed by a permutation test of multivariate homogeneity of group dispersions using the 'permutest' function of the Vegan package. Boxplots were made to analyze the differences in process abundances across different depth profiles. The Kruskal-Wallis test for significance was used to determine whether the overall changes were significant, whereas the Wilcoxon test was used to find the pairwise significance. Heatmap and cluster analyses were carried out with the pheatmap package (Kolde, 2019) and based on normalized gene coverage. Column clustering, which displayed the clustering of different samples, was based on Bray–Curtis dissimilatory. Each row, depicting normalized gene coverage, was scaled using min-max scaling and clustered based on correlation. To understand the clustering of MAGs based on genomic repertoire, genes obtained from Ghost Koala analysis were mapped to each dereplicated MAGs using custom R scripts, and Non-Metric

Multidimensional Scaling (NMDS) of a binary matrix based on the presence or absence of all genes constraining different metabolic categories were performed using Vegan. All the statistical analyses for the MAGs were conducted based on dereplicated MAGs, except for the mapped read coverage analysis. For mapped read coverage, all the MAGs were considered. A web-based tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) was used to generate Venn diagrams to find the overlap across different metabolic categories to determine metabolic lifestyle of the MAGs.

3. Results

3.1.Microbial community function

Microbial community functions varied across depths (Figure 2). Three distinct clusters were observed for deep, medium-deep, and shallow samples with minor overlaps based on normalized gene coverage (PERMANOVA, $p = 0.001$). One of the medium-deep samples (Armstrong Reef) clustered with the shallower-depth samples. Dispersion among sample groups differed according to depth ($p = 0.001$). The average distance from the median was highest for shallower depth samples, followed by medium-deep and deeper samples. The samples were also significantly different (PERMANOVA, $p = 0.011$) based on region (Southern, Northern, and Palmer Canyon). However, the separation of samples was more pronounced for depth than region according to the PERMANOVA F-ratio ($F_{depth} = 10.224$, and $F_{region} = 2.6985$). We conducted a detailed study of carbon, nitrogen, sulfur, and carbohydrate metabolism-specific genes based on normalized gene coverage to understand the differential abundances of functional genes across different depth horizons and regions.

3.1.1. Carbon cycle

genes involved in prokaryotic dark carbon fixation, carbon monoxide oxidation, fermentation, methane oxidation, and photoheterotrophy were explored to understand carbon transformation and their distribution in the wAP. Varied abundances of groups of genes involved in different carbon transformation processes were observed across different depths (Supplementary Figure S1). Genes associated with dark carbon fixation were found to be significantly higher in the deeper samples compared to medium-depth ($p = 9 \times 10^{-7}$) and shallower samples ($p = 0.00085$) (Figure 3A). The abundance of the genes involved in CO oxidation varied across depths. The deeper samples harbored significantly greater CO oxidation gene coverages compared to shallower ($p = 0.00096$)

and medium-deep samples ($p = 4.4 \times 10^{-5}$) (Figure 3B). A similar trend was observed for genes involved in fermentation and methane metabolism (Figures 3C and 3D).

3.1.2. Sulfur cycle

The abundance of genes involved in the oxidation and reduction of sulfur species varied across depths (Supplementary Figure S2). Normalized coverages of genes involved in sulfur oxidation were found to be significantly higher in the shallower samples compared to medium-depth ($p = 0.0051$) and deeper samples ($p = 3.9 \times 10^{-5}$) (Figure 4A). On the contrary, normalized coverage of genes involved in dissimilatory sulfate reduction were found to be significantly higher in the deeper samples compared to medium-depth ($p = 9 \times 10^{-7}$) and shallower samples ($p = 6.9 \times 10^{-6}$) (Figure 4B). Normalized coverage of genes involved in thiosulfate to sulfide reduction was found to be the highest in the deeper samples, followed by medium-depth and shallower samples (Figure 4C).

3.1.3. Nitrogen cycle

The abundances of the genes involved in oxidation and reduction of nitrogen species also varied with depth (Supplementary Figure S3). The average coverages of the genes involved in nitrification were significantly higher in the deeper horizons of the water column compared to medium-depth ($p = 0.0051$) and shallower horizons ($p = 4.1 \times 10^{-7}$) (Figure 5A). Similarly, the abundance of genes involved in denitrification and dissimilatory nitrate reduction to ammonia (DNRA) was found to be significantly higher in deeper samples compared to medium-deep (Denitrification: $p = 5.4 \times 10^{-6}$; DNRA: $p = 0.00012$) and shallower depth (Denitrification: $p = 3.9 \times 10^{-9}$; DNRA: $p = 0.026$) samples (Figures 5B and 5C). Genes involved in urea utilization were also observed in all the wAP samples. Urea-utilizing gene coverages were found to be significantly higher in the deeper horizon samples compared to samples from the medium depth ($p = 4.5 \times 10^{-7}$) and shallower depth ($p = 0.00013$) horizons (Figure 5D).

3.1.4. Carbohydrate transformations

We further analyzed genes involved in the transformation of glycoprotein, cellulose, chitin, pectin, starch, xylans, and xyloglucans to understand the carbohydrate pool and transformation capabilities of microorganisms across different depth horizons of the wAP (Supplementary Figure S4). Cellulose, pectin (RGI), starch, and xyloglucans metabolizing gene coverages were found to

be significantly higher in the shallower depth samples compared to medium-deep (Cellulose: $p = 0.00033$; pectin [RGI]: $p = 8.6 \times 10^{-5}$; starch: $p = 0.0095$; and xyloglucans: $p = 0.0051$) and deeper samples (cellulose: $p = 2.7 \times 10^{-5}$; pectin [RGI]: $p = 1.8 \times 10^{-6}$; starch: $p = 0.00074$; and xyloglucans: $p = 2.3 \times 10^{-5}$) (Figures 6 A,B,C,D). Abundances of chitin metabolizing genes were significantly higher in deeper samples compared to the samples from medium depth ($p = 8.7 \times 10^{-5}$) and shallow samples ($p = 0.0029$) (Figure 6E).

3.2.Distribution, taxonomy, and metabolic profiles of MAGs

A total of 2940 bins were obtained from 48 samples. Of these, 612 bins with genome completeness of more than 70% and contamination of less than 5% were considered MAGs and used for further analysis. These 612 MAGs were filtered down to 609 MAGs by dRep based on genome quality and were dereplicated to a final set of 137 MAGs. 137 MAGs (representing 609 MAGs) covered 13.11%, 5.96%, and 2.12% of the total filtered reads from deep, medium-deep, and shallow samples, respectively (Figure 7). The most abundant MAG found in the shallower waters was affiliated with Bacteroidota (represented by ANT-68) (Figure 7A). This MAG was exclusively observed in the shallow environments of the coastal wAP. Among these 137 MAGs, 64 MAGs were unique to deep samples, whereas 19 and 15 MAGs were unique to the medium and shallow horizons, respectively. 11 MAGs were found in all three depth horizons, and the remaining 28 MAGs were found in two of the three horizons (medium and deep: 18, shallow and medium: 9, and shallow and deep: 1). When sorted by region, 33 MAGs represented MAGs unique to the Southern region, whereas 32 and 4 MAGs represented MAGs unique to the Palmer Canyon and Northern region, respectively. 21 MAGs were found in all three regions, and the remaining 47 MAGs were found in two of the three regions (Southern – Palmer Canyon: 30, Southern – Northern: 14, and Northern – Palmer Canyon: 3). The most abundant MAG found in the medium-depth waters was affiliated with family SAR324 (represented by ANT-96) (Figure 7B). This MAG was observed in the medium-depth and deeper samples. ANT-96 was also found to be the most abundant MAG in deeper waters (Figure 7C). Considering the proportion of reads that mapped to each MAG, the overall highest average mapped read percentage was also found in ANT-96 (2.11% ± 1.22%; representing 7 MAGs from different samples). Detailed genome statistics of the 137 MAGs are present in Supplementary Table S2.

The MAGs were taxonomically diverse. Of 137 MAGs, 10 MAGs were affiliated with the domain Archaea, and 127 MAGs were affiliated with Bacteria. Average nucleotide identity (ANI) or relative evolutionary divergence (RED) values were analyzed for each MAG based on reference genomes from GTDB using GTDB-Tk. RED values were calculated when the MAGs were unable to be classified based on ANI. ANI values greater than 0.95 were obtained for 40 MAGs (2 archaeal MAGs and 38 bacterial MAGs), whereas RED values (ranging between 0.660 to 0.998) were obtained for the rest of the MAGs (Supplementary Table S3). Nine of the archaeal MAGs were affiliated with the phylum Thermoplasmatota whereas one was affiliated with Thermoproteota (Figure 8A). Among the ten archaeal MAGs, two MAGs had formal taxonomic nomeclature (as indicated in International Code of Nomenclature of Prokaryotes) at the genus level (*Nitrosopumilus* and *Thalassarchaeum*), and one of the Thermoplasmatota MAGs had a separate branch from the root of the phylogenomic tree. Bacterial MAGs were assigned to 13 different phyla (Figure 8B). The highest number of MAGs were affiliated with Proteobacteria (53 MAGs), followed by 17 and 14 MAGs affiliated with Planctomycetota and Verrucomicrobiota, respectively. Out of 127 bacterial MAGs, 61, 28, and 5 MAGs had a formal taxonomic nomeclature at the family, genus, and species levels, respectively. Two bacterial MAGs affiliated with class Alphaproteobacteria (ANT-120) and phylum Planctomycetota (ANT-49) had RED values lower than 0.70. Detailed taxonomy classifications of all 137 MAGs are present in Supplementary Table S3. It was interesting to note that one of the MAGs classified as Myxococcota by GTDB-Tk clustered with Proteobacterial MAGs on the phylogenomic tree (Figure 8B).

Distinct groups of MAGs were observed when clustered according to their metabolic profiles. The NMDS plot indicated that the metabolic profiles of bacterial populations were phylum-specific, with minor overlaps among phyla (Figure 9). Distinct clusters for archaeal and bacterial MAGs were observed on the PCoA plot based on genomic repertoire (PERMANOVA, $p = 0.001$) (Supplementary Figure S5). Specific gene sets and pathways were studied to determine the metabolic capabilities of the MAGs (Supplementary Table S4 and Supplementary Figure S6).

3.2.1. Role of MAGs in carbon transformation

Capabilities of glucose utilization and carbon fixation were analyzed based on different pathways. Since these pathways have multiple enzymes involved in them, pathways having ≥70%

completeness in a MAG were considered as the presence of the pathway in the MAG. The details of the pathways and the completeness profiles are described in Supplementary Table S4. The capability to use externally fixed carbon was studied based on the presence of the Embden Meyerhof and Entner–Doudoroff pathways. These two pathways help in the conversion of glucose into pyruvate. 75 MAGs covering ten phyla had the capability of performing the Embden Meyerhof pathway, whereas 42 MAGs covering seven phyla had genomic repertoires for the Entner–Doudoroff pathway. The pentose phosphate pathway was observed in 88 MAGs covering 11 phyla. The citrate cycle (Krebs cycle) was observed in 112 MAGs, whereas the glyoxylate cycle was observed in 40 MAGs. Three carbon fixation pathways *viz.* 3-Hydroxypropionate bicycle, Arnon–Buchanan cycle (reductive citrate cycle), and Calvin cycle (reductive pentose phosphate cycle) were observed among the MAGs. Capabilities of the Arnon–Buchanan cycle were present in a higher number of MAGs (56 MAGs) compared to the Calvin cycle (present in 26 MAGs) and 3-Hydroxypropionate bicycle (present in 8 MAGs). 3-Hydroxypropionate bicycle was only present in Proteobacterial MAGs, whereas Arnon–Buchanan cycle was distributed over MAGs affiliated to Proteobacteria, Verrucomicrobiota, Actinobacteriota, Planctomycetota, Chloroflexota, Thermoplasmatota, SAR324, and Latescibacterota. Calvin cycle was present in MAGs affiliated to Proteobacteria, Actinobacteriota, Chloroflexota, Bacteroidota, Planctomycetota, Gemmatimonadota, and SAR324.

The ability to produce or catabolize small-chain fatty acids and alcohols was studied based on the presence of certain genes in the genomic inventories of the MAGs. Capabilities of alcohol production (EC 1.1.1.1) were found in 53 MAGs, with the majority of the MAGs affiliated with Proteobacteria. The presence of genes encoding for phosphate acetyltransferase (EC 2.3.1.8) and/or acetate kinase (EC 2.7.2.1) was studied to understand acetate metabolism in the MAGs. Genes encoding for acetate metabolizing enzymes were found in 21 MAGs which were affiliated with six different phyla (Proteobacteria, Verrucomicrobiota, Planctomycetota, Actinobacteriota, Latescibacterota, and Bacteroidota). Genes encoding for L-lactate metabolizing enzyme (L-lactate dehydrogenase) were found in 19 MAGs, which were affiliated with the Proteobacteria, Actinobacteriota, Verrucomicrobiota, and Planctomycetota, whereas genes encoding for D-lactate metabolizing enzyme (D-lactate dehydrogenase) were found in five MAGs affiliated with Proteobacteria, Verrucomicrobiota, and Bacteroidota. The gene encoding for propionate

metabolism (propionate CoA transferase) was found in three MAGs affiliated with two different phyla (Proteobacteria and Verrucomicrobiota).

Genes encoding for carbohydrate-active enzymes were studied to understand the polysaccharide-degrading capabilities of the MAGs. Genes encoding for enzymes involved in breaking down amorphous cellulose were found in 30 MAGs affiliated with seven different phyla (Proteobacteria, Planctomycetota, Bacteroidota, Myxococcota, Verrucomicrobiota, Actinobacteriota, and Latescibacterota), whereas genes encoding for the enzyme involved in breaking down crystalline cellulose was found in 11 MAGs covering five phyla (Planctomycetota, Proteobacteria, Verrucomicrobiota, Actinobacteriota, and Myxococcota). Genes coding for chitin-degrading enzymes were found in 85 MAGs (found in all phyla detected in this study except for MAGs affiliated with Thermoproteota, Chloroflexota, and Thermoplasmatota), whereas genes for starch-degrading enzymes were found in 15 MAGs having affiliation with five phyla (Proteobacteria, Verrucomicrobiota, Planctomycetota, Actinobacteriota, and Latescibacterota). Genes encoding enzymes that can perform Xylan and xyloglucan (major components of hemicellulose) degradation were found in 23 and 40 MAGs, respectively. MAGs with the ability to degrade xylan were affiliated with Proteobacteria, Bacteroidota, Myxococcota, Planctomycetota, Actinobacteriota, and Latescibacterota, whereas MAGs with the ability to degrade xyloglucan were affiliated to Proteobacteria, Bacteroidota, Verrucomicrobiota, Planctomycetota, Myxococcota, Latescibacterota, and Marinisomatota. Genes encoding pectin degradation enzymes were found in 31 MAGs (having affiliation with Proteobacteria, Verrucomicrobiota, Planctomycetota, Bacteroidota, Latescibacterota, Marinisomatota, Gemmatimonadota, Acidobacteriota, and Chloroflexota).

3.2.2. Role of MAGs in sulfur and nitrogen transformation

Genes involved in thiosulfate oxidation were found in 26 MAGs having affiliations with five phyla (Proteobacteria, Gemmatimonadota, Acidobacteriota, and SAR324) (Supplementary Figure S6). Three proteobacterial MAGs were found to harbor genes for dissimilatory sulfate reduction.

Genes involved in ammonia oxidation were found in two proteobacterial MAGs and in one archaeal MAG (affiliated to Thermoproteota). Genes involved in the conversion of nitrite to nitrate (one of the steps in nitrification) were found in two proteobacterial MAGs. Genes involved in the conversion of nitrate to nitrite (one of the steps in denitrification or dissimilatory nitrate reduction

to ammonia) were found in two proteobacterial MAGs. Enzymes involved in the conversion of nitrite to nitric oxide (a key step in denitrification) were found in 14 different MAGs covering eight phyla (Proteobacteria, Actinobacteriota, Thermoproteota, Gemmatimonadota, Verrucomicrobiota, Acidobacteriota, Bacteroidota, and Nitrospinota), whereas enzymes involved in the conversion of nitric oxide to nitrous oxide (an additional step in denitrification) were found in two proteobacterial MAGs.

3.3.Putative metabolic lifestyles of MAGs

Overlaps across these categories are reported in Figures 8 and 10A. It was found that the average genome sizes of the mixotrophs were significantly higher compared to heterotrophs ($p = 0.0001$) and autotrophs ($p = 0.02$) (Figure 10B). In the shallower depth samples, a significantly high mapped read percentage was observed in the mixotrophic MAGs (n = 37 MAGs), while compared to the autotrophic (n = 2 MAGs) and heterotrophic MAGs (n = 52 MAGs) (mixotroph vs. autotrophs, $p = 0.054$; mixotroph vs. heterotroph, $p = 1.5 \times 10^{-5}$) (Figure 10C). In the medium and deeper samples, the mapped read percentages for mixotrophic, autotrophic, and heterotrophic MAGs were not significantly different.

3.3.1. Putative mixotrophs

Four categories of mixotrophs covering 52 dereplicated MAGs were obtained (Supplementary Figure S6 and Supplementary Table S4). Only one MAG (ANT-6) recovered from a deeper horizon sample and classified as *Paraburkholderia* was found to harbor genes from all four categories of metabolism (GU, CF, SO, and NI). There were 12 MAGs that harbored genes from GU, CF, and SO (11 affiliated to Proteobacteria and one affiliated to Gemmatimonadota), whereas there was only one MAG that harbored genes from GU, CF, and NI (affiliated to Proteobacteria). There were 38 MAGs that harbored the genes from GU and CF metabolism categories only (affiliated to Proteobacteria, Actinobacteriota, Verrucomicrobiota, Planctomycetota, Chloroflexota, Bacteroidota, Latescibacterota, and SAR324).

3.3.2. Putative autotrophs

Three categories of autotrophs covering 20 dereplicated MAGs were obtained (Supplementary Table S4). There were three MAGs harboring genes for CF and SO (affiliated with Gammaproteobacteria), whereas there was only one MAG harboring gene for CF and NI (affiliated

with Gammaproteobacteria). There were 16 MAGs that had genomic repertoire for CF but lacked genes for SO and NI (affiliated with Proteobacteria, Thermoplasmatota, Planctomycetota, Chloroflexota, Actinobacteriota, and Verrucomicrobiota).

3.3.3. Putative heterotrophs

Two categories of heterotrophs covering 35 dereplicated MAGs were observed in this study (Supplementary Figure S6 and Supplementary Table S4). One of the categories harbored genes for GU and SO, which was found in five different MAGs (affiliated to Proteobacteria and SAR324). The other category solely harbored genes for GU and was found in 30 MAGs (affiliated to Proteobacteria, Bacteroidota, Verrucomicrobiota, Planctomycetota, Actinobacteriota, SAR324, and Myxococcota).

4. Discussion

Our analysis reveals the diverse genomic repertoire contained among marine bacteria and archaea along the coastal wAP. PCoA based on normalized gene coverage across 48 samples suggested that microbial community functions were strongly partitioned by depth, with the highest variation in microbial community function observed in the shallower samples (0-40 m). A possible explanation is that this higher variance results from a more dynamic environment in the shallower waters than in the deeper environments.

Primary production in the coastal Antarctic is attributed primarily to phytoplankton (Arrigo et al., 2008). However, prokaryotic dark carbon fixation can be significant in deep and polar oceans (Alonso-Sáez et al., 2010; Williams et al., 2012; Connelly et al., 2014), and previous work has identified the genomic signatures of dark carbon fixation along the wAP (Grzymski et al., 2012). We observed genes indicative of dark carbon fixation in numerous MAGs representing multiple phyla. This suggests that in addition to eukaryotic carbon fixation, prokaryotic dark carbon fixation can also be a source of fixed carbon in the Antarctic marine ecosystem, expanding the conventional viewpoint of marine primary production (Buchan et al., 2014). We are not aware of any studies that attempt to quantify dark carbon fixation inputs to coastal Antarctic ecosystems. However, previous work suggests oceanic primary production estimates would increase by 5 %–22 % when total dark dissolved organic carbon fixation is included (Baltar and Herndl, 2019). Moreover, the normalized coverage of genes related to dark carbon fixation was found to be higher in deeper

waters compared to the waters from medium-deep and shallower horizons in the coastal wAP suggesting that the dearth of phytoplankton-fixed carbon in the deeper waters selects for microorganisms capable of fixing carbon in the dark. This lack of photosynthate as an electron donor in deeper water is further supported by the presence of higher normalized gene coverage of CO oxidation genes. Lappan et al. (2023) previously reported an enrichment of CO hydrogenase (an enzyme involved in CO oxidation) in temperate mesopelagic waters and suggested that CO oxidation is favored in energy-limited waters at depths where primary production is low.

The normalized gene coverages for processes preferred in hypoxic/anoxic environments, such as fermentation, dissimilatory sulfate reduction, thiosulfate to sulfide reduction, denitrification, and DNRA, were found to be higher in the water samples from the deeper horizons (>100 m). Although oxygen drawdown is observed below the photic zone in the wAP (Cape et al., 2019), the water column typically remains sufficiently oxic to support aerobic processes, including nitrification and methane oxidation in deeper waters. We suggest two explanations for the presence of anaerobic pathways here. First, the organisms harboring these pathways may be facultative anaerobes that can dwell in multiple environments of the wAP. This is supported by the presence of certain MAGs capable of using oxygen (presence of cytochrome c oxidase and F-Type ATPases genes) and nitrate (presence of genes involved in denitrification) as terminal electron acceptors. Second, these microbial populations may come from fecal pellets and sinking detritus that harbor microenvironments that support the growth of anaerobic microbial populations.

Due to high levels of primary production, the coastal wAP harbors a diverse and abundant carbohydrate pool. The source of these carbohydrates includes phytoplankton or zooplankton such as (Oijen et al., 2003; Yu et al., 2020). We observed a diverse genomic repertoire for degrading carbohydrates derived from phytoplankton and zooplankton. This suggests a link between the microeukaryotic and prokaryotic populations in the Antarctic. Moreover, higher abundances of microorganisms capable of degrading cellulose, xyloglucans, pectin, and starch were observed in the upper water column, consistent with the greater phytoplankton biomass expected there. This distribution suggests that carbohydrates are readily turned over in the upper water column, possibly reducing vertical export. This is further supported by the presence of prokaryotic dark carbon fixing pathways in the deeper environment of coastal wAP. Several MAGs having the capability to degrade polysaccharide was also observed in the medium-deep and

deeper horizon of the wAP. This suggests that these microbial populations may be associated with the sinking detritus material, which helps in the conversion of complex carbohydrates to simpler carbon compounds that can be used by other heterotrophs.

Varied metabolic flexibility was observed in several MAGs. ANT-6, affiliated to *Paraburkholderia fungorum* was found to have the most diverse genomic repertoire based on the presence of heterotrophic and autotrophic pathways. The genomic repertoire suggests that they can perform carbon fixation and also utilize glucose. The mixotrophic behavior of *Paraburkholderia* has been reported previously (Herpell et al., 2020). Interestingly, ANT-6 also harbored genes involved in thiosulfate oxidation, nitrification, and denitrification. Sulfite oxidation was previously observed in *Paraburkholderia caledonica* PHRS4 (BioCyc ID: PWY-5276). Even though varied metabolic traits of the genus *Paraburkholderia* are well known, the presence of genomic machinery related to autotrophic and heterotrophic lifestyle, along with the presence of thiosulfate oxidation, nitrification, and denitrification capability in a single strain, is not reported elsewhere to the best of our knowledge. The genome size of ANT-6 was found to be the highest (9.3 Mbp) among all the other MAGs constructed. The genome size of ANT-6 is in the range of genome sizes of other *Paraburkholderia* strains (Wang et al., 2021), which is bigger than "typical" bacterial genomes (Land et al., 2015). The metabolic flexibility of *Paraburkholderia* is consistent with its bigger genomes size. Previous reports of the presence of *Paraburkholderia* in the Antarctic environment (Malard et al., 2022) also support the presence of ANT-6 in the Antarctic.

A high number of mixotrophic MAGs (52 MAGs) were observed compared to exclusively heterotrophic (35) and autotrophic (20) MAGs. The prevalence of mixotrophy was also previously reported in Arctic prokaryotic genomes (Royo-Llonch et al., 2021). In the shallow samples of the coastal wAP, the abundances of mixotrophic MAGs were found to be significantly higher compared to heterotrophic and autotrophic MAGs, whereas, in the deeper samples, the abundances of the mixotrophic MAGs were not significantly different compared to autotrophic and heterotrophic MAGs. This suggests a strong pressure for metabolic flexibility, potentially a response to the seasonal boom-bust cycle of photosynthetic primary production. This is in line with the higher variations of genomic repertoire in the shallower samples compared to medium-depth and deeper samples. We hypothesize that dynamic environments select microbial populations with a diverse genomic repertoire, including the capacity to switch between

heterotrophic and autotrophic lifestyles. Moreover, the genome size of the mixotrophs was found to be significantly larger compared to the autotrophs and heterotrophs. This is in accordance with a previous study where they reported larger genome sizes of the generalist compared to the specialist (Sriswasdi et al., 2017). However, it should be taken into account that higher completeness of MAGs is required to ascertain a particular function or metabolic lifestyle with greater confidence.

Several of our MAGs were only distantly related to the genomes of type strains. A MAG classified as Myxococcota (MAG name here) by GTDB-Tk clustered with Proteobacterial MAGs on the phylogenomic tree (Figure 8B) which might be due its novel taxonomy. There might also be two other reasons for this: (i) lower completeness of the Myxococcota genome (75.54 % completeness) and/or (ii) close relatedness of Myxococcota genome with Proteobacterial genomes, which can be supported by recent classification of Myxococcota as a new separate phylum, which was earlier assigned to class Deltaproteobacteria (Murphy et al., 2021). Only 40 MAGs had ANI values higher than 0.95 when compared to the genomes from the GTDB reference database. Based on our current analyses, we found that only 70, 30, and 5 MAGs had a formal taxonomic nomenclature at the family, genus, and species levels, respectively. We were unable to determine the metabolic lifestyle of the MAG (ANT-120, affiliated to Alphaproteobactria) having the lowest RED value (0.66). Similarly, there were 29 other MAGs for which we were unable to determine the metabolic lifestyle based on the criteria we used in this study. Though this might be a limitation imposed by genome completeness, there are chances of the presence of novel metabolic pathways in them.

This study used high-throughput metagenomics to understand the microbial role in the marine ecosystem of the wAP. A streamlined metagenomic sequence analysis pipeline (iMAGine) was developed to process data and reconstruct MAGs. Our pipeline enabled a coverage-based approach to understand how genes were partitioned by depth and region. With this approach, we identified diverse groups of microorganisms contributing to the carbon, sulfur, and nitrogen cycle along the coastal wAP. Distinct microbial metabolisms were observed across different depth horizons. In particular, higher abundances of mixotrophic MAGs compared to heterotrophic and autotrophic MAGs were found in the shallower waters, suggesting that the dynamic pelagic environment of the coastal wAP has selected microbial populations which can adapt to rapidly changing nutrient availability. Metabolic profiles of the MAGs were phylum specific, indicating a strong link

between functional guilds and taxonomy. Our results highlight the novel genetic and metabolic diversity present within Antarctic marine ecosystems and the need for future studies based on cultivable microbes to better understand the distribution of phenotypic and genotypic traits.

Data availability

Sequence data for metagenomics reads from 48 samples, along with 137 dereplicated MAGs, can be found under NCBI BioProject ID PRJNA894514. The MAGs have been deposited at GenBank under the accession JAPKAB000000000- JAPKFH000000000.

Author Contributions

AD and JB designed the study and developed the first draft of the manuscript. AD developed the iMAGine pipeline and conducted the analysis with assistance from JB. EC, RT, NE, and SD collected the samples and edited the manuscript. EC conducted DNA extractions and processed samples for sequencing. HD, DS, and OS contributed to the study design and writing.

Acknowledgment

We would like to thank the staff of Palmer Station and the crew of *ARSV Laurence Gould* for their assistance. This work was funded by NSF-OPP 1846837 to JSB, NSF-OPP 1440435 to HD, and NSF-OPP 2023425 to OS. JSB was partly supported by a Simons Foundation Early Career Marine Microbial Ecology and Evolution fellowship. NE was funded through a SENECYT graduate fellowship. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929).

Reference

Alonso-Sáez, L., Galand, P. E., Casamayor, E. O., Pedrós-Alió, C., and Bertilsson, S. (2010). High bicarbonate assimilation in the dark by Arctic bacteria. *ISME Journal* 4, 1581–1590. doi: 10.1038/ismej.2010.69.

Arrigo, K. R., van Dijken, G. L., and Bushinsky, S. (2008). Primary production in the Southern Ocean, 1997-2006. *J Geophys Res Oceans* 113. doi: 10.1029/2007JC004551.

Baltar, F., and Herndl, G. J. (2019). Ideas and perspectives: Is dark carbon fixation relevant for oceanic primary production estimates? *Biogeosciences* 16, 3793–3799. doi: 10.5194/bg-16-3793-2019.

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., et al. (2018). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences* 115, E400–E408.

Bowman, J. S., Amaral-Zettler, L. A., Rich, J. J., Luria, C. M., and Ducklow, H. W. (2017). Bacterial community segmentation facilitates the prediction of ecosystem function along the coast of the western Antarctic Peninsula. *ISME Journal* 11, 1460–1471. doi: 10.1038/ismej.2016.204.

Bowman, J. S., and Deming, J. W. (2017). Wind-driven distribution of bacteria in coastal Antarctica: evidence from the Ross Sea region. *Polar Biol* 40, 25–35. doi: 10.1007/s00300-016-1921-2.

Bowman, J. S., Kavanaugh, M. T., Doney, S. C., and Ducklow, H. W. (2018). Recurrent seascape units identify key ecological processes along the western Antarctic Peninsula. *Glob Chang Biol* 24, 3065–3078. doi: 10.1111/gcb.14161.

Bowman, J. S., van Mooy, B. A. S., Lowenstein, D. P., Fredricks, H. F., Hansel, C. M., Gast, R., et al. (2021). Whole Community Metatranscriptomes and Lipidomes Reveal Diverse Responses Among Antarctic Phytoplankton to Changing Ice Conditions. *Front Mar Sci* 8. doi: 10.3389/fmars.2021.593566.

Bowman, J. S., Vick-Majors, T. J., Morgan-Kiss, R., Takacs-Vesbach, C., Ducklow, H. W., and Priscu, J. C. (2016). Microbial Community Dynamics in Two Polar Extremes: The Lakes of the McMurdo Dry Valleys and the West Antarctic Peninsula Marine Ecosystem. *Bioscience* 66, 829–847. doi: 10.1093/biosci/biw103.

Buchan, A., LeCleir, G. R., Gulvik, C. A., and González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* 12, 686–698. doi: 10.1038/nrmicro3326.

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12, 59–60. doi: 10.1038/nmeth.3176.

Bushnell, B. (2014). BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Cantalapiedra, C. P., Herṇandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* 38, 5825–5829. doi: 10.1093/molbev/msab293.

Cape, M. R., Vernet, M., Pettit, E. C., Wellner, J., Truffer, M., Akie, G., et al. (2019). Circumpolar deep water impacts glacial meltwater export and coastal biogeochemical cycling along the west Antarctic Peninsula. *Front Mar Sci* 6. doi: 10.3389/fmars.2019.00144.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* (Oxford University Press), i884–i890. doi: 10.1093/bioinformatics/bty560.

Clarke, A., Murphy, E. J., Meredith, M. P., King, J. C., Peck, L. S., Barnes, D. K. A., et al. (2007). Climate change and the marine ecosystem of the western Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 149–166. doi: 10.1098/rstb.2006.1958.

Connelly, T. L., Baer, S. E., Cooper, J. T., Bronk, D. A., and Wawrik, B. (2014). Urea uptake and carbon fixation by marine pelagic bacteria and archaea during the Arctic summer and winter seasons. *Appl Environ Microbiol* 80, 6013–6022.

Dutta, A., Goldman, T., Keating, J., Burke, E., Williamson, N., Dirmeier, R., et al. (2022). Machine Learning Predicts Biogeochemistry from Microbial Community Structure in a Complex Model System. Available at: https://journals.asm.org/journal/spectrum.

Green, J. L., Bohannan, B. J. M., and Whitaker, R. J. (2008). Microbial biogeography: From taxonomy to traits. *Science (1979)* 320, 1039–1043. doi: 10.1126/science.1153475.

Grzymski, J. J., Riesenfeld, C. S., Williams, T. J., Dussaq, A. M., Ducklow, H., Erickson, M., et al. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *ISME Journal* 6, 1901–1915. doi: 10.1038/ismej.2012.31.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075.

Herpell, J. B., Schindler, F., Bejtović, M., Fragner, L., Diallo, B., Bellaire, A., et al. (2020). The Potato Yam Phyllosphere Ectosymbiont Paraburkholderia sp. Msb3 Is a Potent Growth Promotor in Tomato. *Front Microbiol* 11. doi: 10.3389/fmicb.2020.00581.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47, D309–D314. doi: 10.1093/nar/gky1085.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. Available at: http://www.biomedcentral.com/1471-2105/11/119.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006.

638 Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an
639 adaptive binning algorithm for robust and efficient genome reconstruction from
640 metagenome assemblies. *PeerJ* 7, e7359–e7359. doi: 10.7717/peerj.7359.

641 Kolde, R. (2019). Package 'pheatmap.' Available at: https://cran.r-
642 project.org/web/packages/pheatmap/pheatmap.pdf.

643 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: A
644 fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.
645 *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305.

646 Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., et al. (2015). Insights
647 from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15, 141–161. doi:
648 10.1007/s10142-015-0433-4.

649 Lappan, R., Shelley, G., Islam, Z. F., Leung, P. M., Lockwood, S., Nauer, P. A., et al. (2023).
650 Molecular hydrogen in seawater supports growth of diverse marine bacteria. *Nat Microbiol*.
651 doi: 10.1038/s41564-023-01322-0.

652 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
653 Available at: http://arxiv.org/abs/1303.3997.

654 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence
655 Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:
656 10.1093/bioinformatics/btp352.

657 Lin, Y., Moreno, C., Marchetti, A., Ducklow, H., Schofield, O., Delage, E., et al. (2021). Decline
658 in plankton diversity and carbon flux with reduced sea ice extent along the Western
659 Antarctic Peninsula. *Nat Commun* 12. doi: 10.1038/s41467-021-25235-w.

660 Loy, A., Duller, S., Baranyi, C., Mußmann, M., Ott, J., Sharon, I., et al. (2009). Reverse
661 dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing
662 prokaryotes. *Environ Microbiol* 11, 289–299. doi: 10.1111/j.1462-2920.2008.01760.x.

663 Malard, L. A., Avila-Jimenez, M. L., Schmale, J., Cuthbertson, L., Cockerton, L., and Pearce, D.
664 A. (2022). Aerobiology over the Southern Ocean – Implications for bacterial colonization of
665 Antarctica. *Environ Int* 169. doi: 10.1016/j.envint.2022.107492.

666 McLeod, D. J., Hallegraeff, G. M., Hosie, G. W., and Richardson, A. J. (2012). Climate-driven
667 range expansion of the red-tide dinoflagellate Noctiluca scintillans into the Southern Ocean.
668 *J Plankton Res* 34, 332–337. doi: 10.1093/plankt/fbr112.

669 McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive
670 analysis and graphics of microbiome census data. *PLoS One* 8, e61217.

671 Meredith, M. P., and King, J. C. (2005). Rapid climate change in the ocean west of the Antarctic
672 Peninsula during the second half of the 20th century. *Geophys Res Lett* 32, 1–5. doi:
673 10.1029/2005GL024042.

Murphy, C. L., Yang, R., Decker, T., Cavalliere, C., Andreev, V., Bircher, N., et al. (2021). Genomes of novel myxococcota reveal severely curtailed machineries for predation and cellular differentiation. *Appl Environ Microbiol* 87. doi: 10.1128/AEM.01706-21.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27, 824–834.

Oijen, T. van, Leeuwe, M. A. van, and Gieskes, W. W. C. (2003). Variation of particulate carbohydrate pools over time and depth in a diatom-dominated plankton community at the Antarctic Polar Front. *Polar Biol* 26, 195–201. doi: 10.1007/s00300-002-0456-x.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., et al. (2007). The vegan package. *Community ecology package* 10, 631–637.

Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal* 11, 2864–2868. doi: 10.1038/ismej.2017.126.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25, 1043–1055.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2, 1533–1542. doi: 10.1038/s41564-017-0012-7.

Plum, C., Hillebrand, H., and Moorthi, S. (2020). Krill vs salps: dominance shift from krill to salps is associated with higher dissolved N:P ratios. *Sci Rep* 10, 5911. doi: 10.1038/s41598-020-62829-8.

Royo-Llonch, M., Sánchez, P., Ruiz-González, C., Salazar, G., Pedrós-Alió, C., Sebastián, M., et al. (2021). Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol* 6, 1561–1574. doi: 10.1038/s41564-021-00979-9.

Saba, G. K., Fraser, W. R., Saba, V. S., Iannuzzi, R. A., Coleman, K. E., Doney, S. C., et al. (2014). Winter and spring controls on the summer food web of the coastal West Antarctic Peninsula. *Nat Commun* 5. doi: 10.1038/ncomms5318.

Schofield, O., Brown, M., Kohut, J., Nardelli, S., Saba, G., Waite, N., et al. (2018). Changes in the upper ocean mixed layer and phytoplankton productivity along the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376. doi: 10.1098/rsta.2017.0173.

Seyitmuhammedov, K., Stirling, C. H., Reid, M. R., van Hale, R., Laan, P., Arrigo, K. R., et al. (2022). The distribution of Fe across the shelf of the Western Antarctic Peninsula at the start of the phytoplankton growing season. *Mar Chem* 238. doi: 10.1016/j.marchem.2021.104066.

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., la Rosa, S. L. 0003 3527 8101, Solden, L. M., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48, 8883–8900. doi: 10.1093/nar/gkaa621.

Siegert, M., Atkinson, A., Banwell, A., Brandon, M., Convey, P., Davies, B., et al. (2019). The Antarctic Peninsula under a 1.5°C global warming scenario. *Front Environ Sci* 7. doi: 10.3389/fenvs.2019.00102.

Sriswasdi, S., Yang, C. C., and Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nat Commun* 8. doi: 10.1038/s41467-017-01265-1.

Team, Rs. (2015). RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL http://www. rstudio. com* 42, 84.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi: 10.1038/nature24621.

Wang, K., Wu, Y., Ye, M., Yang, Y., Asiegbu, F. O., Overmyer, K., et al. (2021). Comparative Genomics Reveals Potential Mechanisms of Plant Beneficial Effects of a Novel Bamboo-Endophytic Bacterial Isolate Paraburkholderia sacchari Suichang626. *Front Microbiol* 12. doi: 10.3389/fmicb.2021.686998.

Waters, K. J., and Smith, R. C. (1992). Palmer LTER: A sampling grid for the Palmer LTER program. *Antarct J US* 27, 236–239.

Williams, T. J., Long, E., Evans, F., Demaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME Journal* 6, 1883–1900. doi: 10.1038/ismej.2012.28.

Yu, Y., Liu, X., Miao, J., and Leng, K. (2020). Chitin from Antarctic krill shell: Eco-preparation, detection, and characterization. *Int J Biol Macromol* 164, 4125–4137. doi: 10.1016/j.ijbiomac.2020.08.244.

Figure Legend

Figure 1: Sample locations in the western Antarctic Peninsula. S - only shallower samples analyzed, M- only medium-depth samples analyzed, D - only deeper samples analyzed, S-M-D – samples collected from deep, medium, and shallow environments analyzed.

Figure 2: Principal coordinates analysis (PCoA) of Bray-Curtis dissimilatory based on normalized gene coverage across 48 samples. Symbol shape and color indicates sample region and depth, respectively.

Figure 3: Boxplot showing the distribution of different carbon cycle genes related to (A) carbon fixation, (B) CO oxidation, (C) fermentation, and (D) methane oxidation across different depth horizons of the WAP. Normalized coverages of genes grouped in each of these categories are listed in Supplementary Figure S1. A pairwise comparison for significance was conducted using the Wilcoxon test.

Figure 4: Boxplot showing the distribution of different sulfur cycle genes related to (A) sulfur oxidation, (B) dissimilatory sulfate reduction, and (C) thiosulfate to sulfide reduction across different depth horizons of the WAP. Normalized coverages of genes grouped in each of these categories are listed in Supplementary Figure S2. A pairwise comparison for significance was conducted using the Wilcoxon test.

Figure 5: Boxplot showing the distribution of different nitrogen cycle genes related to (A) nitrification, (B) denitrification, (C) dissimilatory nitrate reduction to ammonia (DNRA), and (D) urea utilization across different depth horizons of the WAP. Normalized coverages of genes grouped in each of these categories are listed in Supplementary Figure S3. A pairwise comparison for significance was conducted using the Wilcoxon test.

Figure 6: Boxplot showing the distribution of different carbohydrate transformation genes related to (A) cellulose, (B) pectin RGI, (C) starch, (D) xyloglucan, and (E) chitin degradation across different depth horizons of the WAP. Normalized coverages of genes grouped in each of these categories are listed in Supplementary Figure S4. A pairwise comparison for significance was conducted using the Wilcoxon test.

Figure 7: rage abundance of 137 dereplicated MAGs across (A) Shallow, (B) Medium, and (C) Deep waters of the coastal wAP. Each dereplicated MAG represents multiple MAGs from different samples. The average abundance of each dereplicated MAG represents the average abundance of all the MAGs it represents from a particular depth horizon.

Figure 8: Maximum-likelihood phylogenomic analysis of 137 dereplicated MAGs based on (A) archaeal and (B) bacterial marker genes from GTDB-Tk analysis. Detailed genome statistics and taxonomy are provided in Supplementary Table S3, and the detailed pathways and functions are provided in Supplementary Table S4. Phylum, completeness %, contamination %, and bootstrap legends are domain-specific, whereas depth, region, and function are common for bacteria and archaea. Depth (represented by a star) and region (represented by a right triangle) symbols are present on the terminal node and terminal branch, respectively. The circle symbols in the outermost layer depict the functional capabilities of each MAG. All the carbohydrates mentioned represent the presence of enzymes capable of degrading them. Xylan and/or xyloglucan degrading enzymes are represented by hemicellulose in this figure. The hyphenated legends for depth represent MAGs found in multiple depth horizons, whereas hyphenated legends for regions represent MAGs found in multiple regions.

Figure 9: Non-Metric Multidimensional Scaling (NMDS) of binary matrix based on the presence or absence of genes predicted by Ghost Koala across 127 dereplicated bacterial MAGs. Abbreviations of metabolic categories as used in the CAP analysis are as follows: GU: Glucose

utilization; CF: carbon fixing pathways; NI: nitrification; SO: sulfur oxidation; DNR – Denitrification; DSR – Dissimilatory sulfite reduction.

Figure 10: Comparison of MAGs having different metabolic lifestyles (A) Venn Diagram showing overlaps among different metabolic categories of the MAGs. GU: Glucose utilization; CF: carbon fixing pathways; NI: nitrification; SO: sulfur oxidation. Boxplot showing genome sizes of MAGs related to different metabolic lifestyles. The details of the genome sizes and metabolic lifestyle of each MAG are reported in Supplementary Table 3. Boxplot showing mapped read percentage of MAGs having different metabolic lifestyles in shallower depth samples. For the boxplots, a pairwise comparison for significance was conducted using the Wilcoxon test.

Supplementary Table Legend

Table S1: Sample metadata for 48 samples

Table S2: Detailed genome statistics of 137 MAGs

Table S3: Detailed taxonomic classification-based genome summary

Table S4: Metabolic profiles of 137 MAGs