Antarctic geothermal soils exhibit an absence of regional habitat generalist microorganisms

Stephen E. Noell,[a,b] Jaber Abbaszadeh,[a,b] Huw Richards,[a,b] Marie Labat Saint Vincent,[a,b,c] Charles K. Lee,[a,b] Craig W. Herbold,[d] Matthew B. Stott,[d] S. Craig Cary[a,b], Ian R. McDonald[a,b]*

[a] Thermophile Research Unit, Te Aka Mātuatua | School of Science, Te Whare Wānanga o Waikato | University of Waikato, Hamilton, Aotearoa-New Zealand

[b] International Centre for Terrestrial Antarctic Research, Te Whare Wānanga o Waikato | University of Waikato, Hamilton, Aotearoa-New Zealand

[c] University de Lille, Lille, France

[d] Te Kura Pūtaiao Koiora | School of Biological Sciences, Te Whare Wānanga o Waitaha | University of Canterbury, Christchurch, Aotearoa-New Zealand

Running title: Antarctic Geothermal Microorganisms

* Address correspondence to Ian R. McDonald:

Te Aka Mātuatua - School of Science, Te Whare Wānanga o Waikato - University of Waikato, Hamilton, New Zealand

Email: ian.mcdonald@waikato.ac.nz

Abstract

Active geothermal systems are relatively rare in Antarctica and represent metaphorical islands ideal to study microbial dispersal. In this study, we tested the macroecological concept that high dispersal rates result in communities being dominated by either habitat generalists or specialists by investigating the microbial communities on four geographically separated geothermal sites on three Antarctic volcanoes (Mts. Erebus, Melbourne and Rittman). We found that the microbial communities at higher temperature (max 65□) sites (Tramway Ridge on Erebus and Rittmann) were unique from each other and were dominated by a variety of novel *Archaea* from class *Nitrososphaeria*, while lower temperature (max 50□) sites (Western Crater on Erebus and Melbourne) had characteristically mesophilic communities (*Planctomycetes*, *Acidobacteriota*, etc) that were highly similar. We found that 97% of the detected microbial taxa were regional habitat specialists, with no generalists, with community assembly driven by high dispersal rates and drift (25 and 30% of community assembly, respectively), not environmental selection. Our results indicate that for microbial communities experiencing high dispersal rates between isolated communities, habitat specialists may tend to out-compete habitat generalists.

Introduction

In ecology, the niche concept is commonly used to group organisms based on the range of different habitats they live in. Habitat generalists (which tend to be metabolic generalists (Barberán *et al.*, 2014; Bell and Bell, 2021; Chen *et al.*, 2021a; Malard and Guisan, 2023)) are adapted to thrive in a broader range of different ecosystems compared to habitat specialists, although some organisms may be specialized to a habitat in one area but have a broader distribution beyond the scope of the current study (Malard and Guisan, 2023). Habitat generalists that have a global distribution are usually called “cosmopolitan” to distinguish them from other habitat generalists (Herbold, Lee, *et al.*, 2014). In most

environments, the diversity of habitat specialists is greater than generalists, but habitat generalists tend to be the most abundant organisms; this is true both for microorganisms (Barberán *et al.*, 2014; Zhang *et al.*, 2018; Chen *et al.*, 2021b; Gong *et al.*, 2022) and macroorganisms (Brown, 1984; Magurran and Henderson, 2003). Exceptions to this trend are few, but in macroecology, high rates of dispersal between communities can lead to habitat specialists outcompeting habitat generalists, although the opposite has also been observed (Brown and Pavlovic, 1992). In contrast, at least two reported exceptions to this trend (habitat generalists being the most abundant) in microbial ecology have been found under situations of severe environmental selection and/or dispersal limitation: strong salinity gradients (Logares *et al.*, 2013) and the mammalian gut (Green *et al.*, 2016). Other environmental factors are well known to be highly important in structuring microbial communities (e.g., pH (Fierer and Jackson, 2006; Rousk *et al.*, 2010), salinity (Jiang *et al.*, 2022), water availability (Niederberger *et al.*, 2019)), raising the possibility that environmental factors may be the most important factor determining the proportion of generalists and specialists in environments. In microbial ecology, it is unknown how high dispersal rates would influence the direction of community dynamics: would it result in habitat specialists finding their optimal niches and outcompeting all generalists, resulting in no generalists? Or, would it lead to an enrichment in habitat generalists that are capable of filling any niche into which they are dispersed, given their strong ability to survive dispersal (Sriswasdi *et al.*, 2017; Schiro *et al.*, 2022)? It is challenging to answer this question, however, given the complexity of most systems (many routes of dispersal, ecological disturbances, the influence of macroorganisms, etc.).

Antarctic geothermal regions offer a unique opportunity to resolve the outcome of high dispersal rates on the proportions of habitat generalists and specialists in a microbial community for several reasons. First, Antarctica is one of the windiest places on earth

(Parish, 1988), which is likely to result in strong dispersal of microorganisms between sites within Antarctica, as has been found in the McMurdo Dry Valleys (Archer *et al.*, 2019) and other areas of Antarctica (Hughes *et al.*, 2004; Pearce *et al.*, 2010). Second, aerial dispersal is the only possible route of dispersal between Antarctic geothermal regions, given the absence of aquifers and bird traffic connecting regions, which reduces complexity. Third, Antarctic geothermal regions have relatively simple communities, with little to no macroscopic life present (Herbold, Mcdonald, *et al.*, 2014). Finally, Antarctic geothermal regions serve as refugia for thermophiles, given the vast amount of ice and snow separating regions and the extremely cold air conditions in Antarctica, resulting in little regional influence of non-thermophiles on the microbial communities present. Interestingly, past studies of thermophiles, mainly in hot spring environments, have found high niche specialization for temperature or pH, with environmental selection being the primary factor influencing community assembly and low dispersal rates between sites (Weltzer and Miller, 2013; Power *et al.*, 2018a, 2023; Mahato *et al.*, 2019; Fernandes□Martins *et al.*, 2023). It is unknown whether the strong winds in Antarctica are capable of overcoming the generally low dispersal rates for thermophiles at Antarctic geothermal regions.

The southernmost active geothermal features in the world are located in Antarctica, surrounded by vast swathes of ice and snow: Mt. Erebus, Mt. Melbourne, Mt. Rittmann, and Deception Island (Herbold, Mcdonald, *et al.*, 2014). Within these geothermal areas, the interplay between cold air (as low as -60□) and hot soils results in a large variety of unique features such as ice caves, ice chimneys, and ice hummocks. Of these regions, Mt. Erebus is the highest (3794m above sea level, a.s.l.) and most volcanically active. Two of its largest geothermal sites are Tramway Ridge and Western Crater, which have differing maximal temperatures (65□ vs 50□), geothermal features (many active fumaroles in bare soil producing large amounts of steam vs. warm soil mostly covered by ice hummocks with less

steam), physicochemistry (e.g., low pH vs. high pH), and productivity levels (visible phototrophic mats and moss beds vs. primarily chemotrophic driven) (Noell *et al.*, 2022). Mt. Melbourne (2733m a.s.l.) is located 358 km northeast of Mt. Erebus and only has two geothermal fields, with maximal temperatures of 50□, limited fumarolic activity, and some moss beds (Broady *et al.*, 1987). Geothermal activity on Mt. Rittmann (2600m a.s.l.), 450 km northeast of Mt. Erebus and 103 km north of Mt. Melbourne, consists of one steep wall with actively steaming fumaroles and hot soil (maximum 65□) that is relatively shallow due to the steep aspect (Bargagli *et al.*, 1996). Given the strong maritime influence, presence of vascular plants, and off-continent location of Deception Island north of the Antarctic circle, we excluded this site from consideration in our study. For this study, we classified the two higher temperature sites with highly active fumarolic activity (Tramway Ridge on Mt. Erebus and Mt. Rittmann) as "Active" sites and the two lower temperature sites with less active fumaroles (Western Crater on Mt. Erebus and Mt. Melbourne) as "Passive" sites.

The microbial communities present at these unique geothermal regions have been incompletely characterized, especially at Mts. Rittmann and Melbourne. On Mt. Erebus, culture-independent microbial community analyses have shown that cosmopolitan, thermophilic microorganisms dominate the surface community of Tramway Ridge, while novel and endemic members of poorly understood archaeal and bacterial groups are dominant >2cm below the surface (Soo *et al.*, 2009; Herbold, Lee, *et al.*, 2014; Herbold *et al.*, 2024). Our study comparing the communities at Tramway Ridge and Western Crater along thermal transects found distinct microbial communities, primarily driven by strong differences in pH and water content between the two sites (Noell *et al.*, 2022). Studies of life on Mt. Melbourne have focused on macroscopic organisms (Broady *et al.*, 1987; Logan *et al.*, 2000; Skotnicki *et al.*, 2001) or were culture-based (Logan *et al.*, 2000; Bargagli *et al.*, 2004; Pepi *et al.*, 2005), with two metagenomic datasets showing a community comprised of *Proteobacteria*,

Chloroflexota, and candidate phylum AD3 (now *Pseudomonadota*, *Chloroflexota*, and *Dormibacteraeota* respectively) (Park *et al.*, 2021; Myeong *et al.*, 2024). Similarly, only culture-dependent studies of the microorganisms at Mt. Rittmann have been conducted to-date (Bargagli *et al.*, 1996; Logan *et al.*, 2000; Allan *et al.*, 2005; Poli *et al.*, 2006). These high-elevation, Antarctic geothermal sites have been theorized to have acted as refugia for terrestrial Antarctic life during highly glaciated periods (Fraser *et al.*, 2014).

In this study, we used a combination of 16S rRNA gene amplicon sequencing and metagenomics to resolve the proportion of habitat generalists at these sites (i.e., do we see a large group of microorganisms common to all sites?), and determine the ecological factors shaping community assembly. Given that the microbial communities at these sites are largely uncharacterized, we began by describing the taxonomy and functionality of the microbes present at these sites before answering our ecological question. We hypothesized that the lower temperature sites (Western Crater on Erebus and Melbourne) would have different microbial communities than the higher temperature sites (Tramway Ridge on Erebus and Rittmann), and that there would be a high proportion of habitat generalists across sites, driven by high rates of aerial dispersal.

Experimental Procedures

Site Description and Sample Collection

Soil samples were collected from sites on Mt. Erebus, Mt. Rittmann, and Mt. Melbourne, Antarctica, between 2010 and 2019. Sampling sites with coordinates are given in Figure 1. On Mt. Erebus, two sites were sampled; the first was Tramway Ridge (TR), which is the largest geothermal site on the Mt. Erebus summit and lies within Antarctic Specially Protected Area (ASPA) No. 175. TR is relatively protected from the wind, has large numbers of actively steaming fumaroles, producing gases enriched in CO_2 and CO (Ilanko *et al.*, 2019), with hot soil reaching 65 in places, and has extensive moss and lichen mats covering

16% of the site (Management Plan For Antarctic Specially Protected Area No . 175, 2014). The second site sampled on Mt. Erebus was Western Crater (WC), which is more exposed to the wind and has very few active fumaroles and no visible moss or lichen mats; soil temperatures only reach 50□ in ice-free grounds. On the Mt. Melbourne summit, samples were collected on Cryptogram Ridge (also protected by ASPA No. 175), which, similar to WC, is comprised of ice-free hot soil spots where temperatures reach 50□. There are some patches of moss on Melbourne, and the few active fumaroles on Melbourne are covered by ice hummocks, generating CO_2 and methane-rich gases (Skotnicki *et al.*, 2001; Management Plan For Antarctic Specially Protected Area No . 175, 2014). On Mt. Rittmann, samples were collected from the single geothermally active area on the mountain, which is also within ASPA No. 175. This site is particularly steep (~40 degree slope), with ice free hot soil reaching 63□ and many actively steaming fumaroles with gases of unknown composition, as well as a few moss patches (Management Plan For Antarctic Specially Protected Area No . 175, 2014).

Soil temperatures were measured prior to sampling using a Checktemp 1C electronic temperature sensor (Hanna Instruments, Rhode Island, United States). Mt. Erebus samples were collected along temperature transects, as detailed in (Noell *et al.*, 2022). Samples were collected using a sterile spatula after first discarding the top centimetre of soil, except for one sample each at Rittmann and Melbourne where the top mat-like layer of soil was taken as the sample (Ritt3M-3 and Mel1M-2). Samples were placed into sterile 50 mL tubes and frozen for transport to the University of Waikato (Hamilton, New Zealand) for analysis.

Geochemistry

Soil samples were analysed for geochemistry using a variety of analytical techniques. pH and electroconductivity (EC) were measured by adding MilliQ water to soil samples at a 1:2.5 soil:water ratio. pH and EC measurements were made using a Thermo Scientific Orio 4-

Star Benchtop pH/Conductivity Meter (Thermo Fisher Scientific) using a 3-point calibration. For gravimetric water content (GWC) measurements, 3 g of soil sample was dried at 105□ until the sample weight remained unchanged. GWC was then calculated as the percentage of weight loss during drying (Herbold, Lee, *et al.*, 2014). For elemental composition (B, Na, Mg, P, K, Ca, V, Cr, Fe, Mn, Co, Ni, Cu, Zn, Ga, As, Se, Sr., Ag, Cd, Ba, Tl, Pb, and U), dried samples (1g) were prepared and analysed as described in (Noell *et al.*, 2022). For total carbon and nitrogen measurements, dried samples were also ground as for elemental analysis, then sent to Analytica (Hamilton, New Zealand) for analysis. For total organic carbon (TOC) and nitrogen (TON) measurements, samples were prepared similarly to (Noell *et al.*, 2022). Briefly, 1g of samples were weighed out into acidified glass jars with cardboard disks in lids; jars and lids were washed as for the ICP-MS protocol. 1 mL 6N HCl was added to samples and dried in an oven in a fume hood for 24 hours at 95□. 100 mg aliquots of samples were analysed for TOC and TON using an automatic CN analyser.

DNA Extraction, Amplification, and Sequencing

For 16S rRNA gene amplification as well as metagenome sequencing, genomic DNA was extracted from 0.5 to 0.9 g of each soil sample using a modified CTAB/bead beating method (Noell *et al.*, 2022). To have enough DNA from low biomass samples for metagenomes, multiple replicates (up to 10) of samples were extracted and pooled. Genomic DNA was quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Massachusetts, United States).

The 16S rRNA gene (V4-V5 region) was amplified using PCR with the 515YF (5’CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXXXXGATGTGYCAG CMGCCGCGGTAA3’) and 926R (5’CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATCCGYCAATTYMTT TRAGTTT3’) primers (Caporaso *et al.*, 2018; Ul-Hasan *et al.*, 2019), adapted for Ion Torrent

sequencing using fusion primers with a unique tag (Whiteley *et al.*, 2012). PCR reaction conditions, library prep, and 16S rRNA gene sequencing are the same as described in Noell, S. E. *et al.* (2022) (Noell *et al.*, 2022). 16S rRNA gene sequencing was conducted on all 21 samples on an Ion Torrent PGM (Thermo Fisher Scientific) at the Thermophile Research Unit (Hamilton, NZ).

Four samples were chosen for shotgun metagenome sequencing: Ritt1-3, Ritt3M-1 (mat sample from Rittmann), WC-50, and Mel1-3; the mat sample was included to improve the odds of recovering cyanobacterial genomes. A fifth sample (from Tramway Ridge) was previously extracted and sequenced using Illumina Hi-Seq; the two surface metagenome samples from this previous project (SRR6519253 and SRR6519256) were downloaded from the SRA using SRAtoolkit v3.0.1 and combined into one larger sequence file. For the four new metagenomic DNA samples, extracted DNA was sent to GENEWIZ Genomics Center (Suzhou, China). Next generation sequencing library preparations were constructed following the manufacturer's protocol (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®). For each sample, 1 μg genomic DNA was randomly fragmented to <500 bp by sonication (Covaris S220). The fragments were treated with End Prep Enzyme Mix for end repairing, 5' Phosphorylation and dA-tailing in one reaction, followed by a T-A ligation to add adaptors to both ends. Size selection of Adaptor-ligated DNA was then performed using AxyPrep Mag PCR Clean-up (Axygen), and fragments of ~410 bp (with the approximate insert size of 350 bp) were recovered. Each sample was then amplified by PCR for 8 cycles using P5 and P7 primers, with both primers carrying sequences which can anneal with flowcell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR products were cleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Then libraries with different indexes

were multiplexed and loaded on an Illumina HiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150 paired-end (PE) configuration; image analysis and base calling were conducted by the HiSeq Control Software (HCS) + RTA 2.7 (Illumina) on the HiSeq instrument.

Sequence Quality Control and Taxonomic Assignment

For 16S rRNA gene amplicon sequences, the raw reads from the Ion Torrent PGM were processed using the DADA2 pipeline with recommended Ion Torrent parameters (Callahan *et al.*, 2016). Chimeras were removed using the function removeBimeraDenovo, method "consensus." Taxonomy assignment was performed using assignSpecies with the naïve Bayesian classifier and SILVA database version 138.1. Eukaryotic, mitochondrial, and chloroplast sequences were removed and an unrooted phylogenetic tree built using the neighbour-joining method with the decipher (Wright, 2016) and phangorn (Schliep, 2011) packages in R, generating final amplicon sequence variants (ASVs).

Unfortunately, two samples (one from Tramway Ridge, TR2-64, one from Western Crater, WC-10) produced no sequence data. As these two samples had also been sequenced in a previous sequencing run (from (Noell *et al.*, 2022)), we investigated whether sequence data for these two samples from the previous run could be safely used with the second sequence run samples. To control for batch effects, we tested whether other samples from the two runs had similar microbial communities and found that there was no pronounced batch effect for all samples between the two runs (Figure S1). Thus, we decided to use sequence data for TR2-64 and WC-10 from the first sequencing run alongside the second run; we denote these samples as TR2-64-1 and WC-10-1 to distinguish them from samples in the second run (TR1-32-2, WC-20-2, etc.).

For metagenome sequences, demultiplexing was performed by bcl2fastq 2.17. Raw data was filtered as follows: (1) Discard pair-end reads with adapter; (2) Discard pair-end

reads when the content of N bases is more than 10% in either read; (3) Discard pair-end reads when the ration of bases of low quality (Q<20) is more than 0.5 in either read. All four metagenome samples were quality checked using FastQC v0.11.5 (Andrews) and the reports visualized using MultiQC (Ewels *et al.*, 2016). Illumina adaptors were trimmed using trimmomatic v0.39 (Bolger *et al.*, 2014) and sequences were trimmed for quality using the following command:

trimmomatic PE -phred33 <input files> ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:2:True MINLEN:70 SLIDINGWINDOW:4:15

In light of the large amount of duplicate sequences observed in some samples (especially the Rittmann sample) and recent evidence that suggests deduplication is beneficial (Zhang *et al.*, 2022), we used fastp v0.23.2 (Chen *et al.*, 2018) to deduplicate all metagenome sequences using the following command:

fastp –in1 <input file forward reads> --out1 <output forward reads> --in2 <input reverse reads> --out2 <output reverse reads> -A -D -G -Q -L –html <summary file>

Reads with ambiguous bases were removed using prinseq-lite v0.20.4 (Schmieder and Edwards, 2011) using the -ns_max_n 0 command.

Metagenome analysis, assembly, and binning

We assessed the taxonomy of the quality trimmed metagenomes at the read level using singleM v0.13.2 (https://github.com/wwood/singlem) with the singleM pipe command; we saw no differences when we used the raw, non-quality trimmed sequences or the quality trimmed sequences (data not shown). We used Nonpareil v3.401 (Rodriguez-R and Konstantinidis, 2014; Rodriguez-R *et al.*, 2018) to assess the fraction of community diversity sampled by the metagenome.

For metagenome assembled genome (MAG) assembly, binning, and processing, we used the metaWRAP pipeline v1.3.2 with default parameters (Uritskiy *et al.*, 2018). We used

Metaspades (Nurk *et al.*, 2017) as the assembly program within metaWRAP with default settings. To minimize the amount of ambiguous bases in the assemblies, we used the non-scaffolded contigs for binning; these non-scaffolded contigs produced bins that were similar in number and quality to the scaffolded contigs (for scaffolded vs non-scaffolded, the bins produced had 94.1% vs 94.0% average completeness, 1.5% vs 1.6% average contamination, 63% vs. 64% average GC, 41309 vs. 40696 average N_{50}). We used prodigal v2.6.3 in metagenome mode (Hyatt *et al.*, 2010, 2012) to predict coding genes from the non-scaffolded assemblies; these predicted coding sequences were used for the metagenome functional analysis. For binning contigs in the metaWRAP pipeline, we used metaBAT2 (Kang *et al.*, 2019), MaxBin2 (Wu *et al.*, 2016), and CONCOCT (Alneberg *et al.*, 2014). For the bin_refinement module of metaWRAP, we used targets of 80% and 5% for completeness and contamination, respectively. Bins were then re-assembled and cleaned with MAGpurify v2.1.2 (Nayfach *et al.*, 2019) using the phylo-markers, clade-markers, gc-content, known-contam, and clean-bin modules to produce the final MAGs. The final statistics for these MAGs were produced using checkM v1.0.12 (Parks *et al.*, 2015) with the lineage_wf command. We used GTDB-Tk v2.1.1 (Chaumeil *et al.*, 2022) for taxonomy assignment of MAGs (GTDB database R214) and both Prodigal and Prokka v1.11 (Seemann, 2014) for predicting coding gene regions. We predicted optimal growth rate of MAGs using gRodon (Weissman *et al.*, 2021) and optimal growth temperature using Tome v1.0.0 (Li *et al.*, 2019). Finally, we used pfam_scan (https://github.com/aziele/pfam_scan) to predict pfam domains in MAGs (pfam release 35.0).

We calculated the presence/absence of MAGs across sites by utilising MinHash sketching techniques implemented with sourmash (Brown and Irber, 2016). Specifically, we created sourmash signatures of all MAGs and quality filtered metagenomic read sets from all four sites using a k-mer size of 21 and scaled to 1000. These settings allowed us to query for

the presence of a MAG within a site with greater sensitivity than by using a higher k-mer or scale value. We selected these metrics to minimize the false negative of not finding a MAG that is present, though we acknowledge this possibly increases the false positive, something recently found by other groups (Koslicki *et al.*, 2023). Finally, we included abundance weighting in the signature creation step of the MAG and metagenome signatures so we could measure relative abundances of MAGs present. To query presence/absence, we used sourmash *gather* on each MAG against each metagenome to calculate containment scores. These scores are a proxy for Average Nucleotide Identity where a containment score of 0.2 is equivalent to an ANI of 0.95 for a signature created with a k-mer of 21. We therefore concluded that a MAG was present within a metagenome if its score was greater than 0.2, to account for the varying completeness levels of the MAGs and to again minimize the false negative rate. We measured MAG abundances within each site using the abundance weighted percent match between the MAG and the metagenome. This correlates to the proportion of metagenomic reads that would map to the MAG, a proxy for abundance. The final set of MAGs was deduplicated for all analyses using a 95% ANI cutoff. Where duplicate MAGs were found, the MAG from the site with its highest abundance was retained.

Analysis of 16S rRNA Gene Sequence Data

Using 16S rRNA gene sequences, microbial community structure was assessed using PcoA of center log-ratio transformed (CLR) sequence abundance on a Unifrac distance matrix using the vegdist function. Statistical differences between sites were evaluated first by testing samples did not differ significantly in their dispersion (using an analysis of multivariate homogeneity, PERMDISP), then by using a permutational analysis of variance (PERMANOVA) within the R package vegan v2.6-2 (Oksanen *et al.*, 2020). Alpha diversity was also assessed within vegan via the Shannon index. Venn diagrams were plotted using the R package ggVennDiagram v1.2.0 (Gao, 2021); a custom modification to the

ggVennDiagram function was used to color the sections of the Venn Diagram by the abundance of ASVs in those sections. Heatmaps were plotted using the R package pheatmap v1.0.12 (Kolde, 2019). The random forest differential abundance analysis was conducted using the R package microeco v0.13.0 with taxa level set to “all” (Liu *et al.*, 2021).

To identify physicochemical parameters that correlated with microbial community structure, the mantel test function within microeco was used with default parameters. The factors that had a non-adjusted two-sided p-value < 0.05 were plotted using a distance-based redundancy analysis (dbRDA). To test for significant differences between physicochemical parameters, the cal_diff (method = anova) and plot_diff functions within the trans_env function of microeco were used. To analyse clustering of sites by physicochemical parameters, principal coordinates analysis (PcoA) of physicochemical parameters in a Euclidian distance matrix was performed using the vegdist function in the microbiome package v.1.12.0 (Lahti and Shetty, 2017).

Metagenome Functional Analysis

We estimated the metabolic potential of each site using a method described by Greening and colleagues (Greening *et al.*, 2019). In short, we used DIAMOND v. 2.1.6 (query cover > 80%) to search quality filtered assembled ORFs of the metagenomes against custom protein databases of representative metabolic marker genes designed by the authors. The metabolic marker genes include multiple pathways or energy harvesting systems, namely: sulfur cycling (AsrA, FCC, Sqr, DsrA, Sor, SoxB), succinate oxidation (SdhA), selenium cycling (YgfK), rhodopsin (RHO), reductive dehalogenation (RdhA), phototrophy (PsaA, PsbA, energy-converting microbial rhodopsin), oxidative phosphorylation (AtpA), nitrogen cycling (AmoA, HzsA, NifH, NarG, NapA, NirS, NirK, NrfA, NosZ, NxrA, NorB), NADH oxidation (NuoF), methane cycling (McrA, MmoA, PmoA), isoprene oxidation (IsoA), iron cycling (Cyc2, MtrB, OmcB), hydrogen cycling (catalytic subunit of [NiFe]-

hydrogenases, the catalytic domain of [FeFe]-hydrogenases, and Fe-hydrogenases), fumarate reduction (FrdA), carbon monoxide oxidation (CoxL, CooS), carbon fixation (RbcL, AcsB, AclB, Mcr, HbsT, HbsC), arsenic cycling (ARO, ArsC), and aerobic respiration (CoxA, CcoN, CyoA, CydA). We filtered results based on different identity thresholds: PsaA (80%), PsbA and IsoA, AtpA, ARO, YgfK (70%), HbsT (75%), group 4 [NiFe]-hydrogenases, [FeFe]-hydrogenases, CoxL, AmoA, NxrA, NuoF (60%), and all others with an identity threshold of 50%. We normalized the relative abundance of present markers to RPKM.

RPKM is recommended for relative abundance comparisons within metagenomic datasets because it normalizes the data based on both sequence depth (per million reads) and sequence length (in kilobases) (Liao *et al.*, 2023). We used a three-step procedure to calculate the relative abundance of each metabolic marker in the community. First, we used quality controlled reads to normalize coverage of ORFs to RPKM using CoverM v0.6.1 contig (https://github.com/wwood/CoverM) (with parameters: -p bwa-mem -min-read-aligned-percent 0.75 -min-read-percent-identity 0.95 --min-covered-fraction 0 -m rpkm). In the second step, we aligned high-quality unassembled reads to 14 universal single-copy ribosomal marker genes used in SingleM v.0.12.1 (https://wwood.github.io/singlem/) and PhyloSift (Darling *et al.*, 2014) using DIAMOND (query cover > 80%, bitscore > 40) and normalized to RPKM as above. Finally, we calculated the average gene copy number of each marker in the community by dividing the read count to the gene (in RPKM) by the mean of the read counts of the 14 universal single-copy ribosomal marker genes (in RPKM). Finally, we visualized the data using ggplot2.

Phylogenomic Tree Analysis

For the *Cyanobacteria* phylogenomic tree, we used GtoTree v1.7.00 (Lee, 2019), with our 9 cyanobacterial MAGs along with GTDB species cluster representatives from the same three Cyanobacterial families, resulting in 217 members of *Nostocaceae*, 38 for

Leptolyngbyaceae, and 55 for *Obscuribacteraceae* (accessed on 13 October 2023). We used the *Cyanobacteria*-specific single copy gene (SCG) set available in the GtoTree package. The SCG threshold was lowered to 35% to keep the *Obscuribacteraceae* genomes. We used four members of the Terrabacteria supergroup (*Streptomyces coelicolor*, *Clostridium botulinum*, *Chloroflexus aurantiacus*, and *Meiothermus ruber*) as the outgroups to root the tree. The tree was visualized and edited using the website Interactive Tree of Life (https://itol.embl.de/upload.cgi).

Community Assembly Analysis

To compare the abundance of cosmopolitan and site-specific microorganisms in the sites/samples, ASVs were categorized by the number of sites they were present at, with the threshold for presence at a site being greater than 0 relative abundance within either the sample or site. MAGs were classified as present or absent as described above.

To calculate niche breadth for ASVs, we first eliminated ASVs only found at one site (1878 ASVs reduced to 832 ASVs), equivalent to a 5% sample occupancy cutoff (stricter than (Sriswasdi *et al.*, 2017; Xu *et al.*, 2022) but not as conservative as (Hernandez *et al.*, 2023)). Next, we used the R package MicroNiche v1.0.0 (Finn *et al.*, 2020), with the default settings for the limit of quantification (LOQ) threshold (q = 1.6). Levin's B_N was used to detect generalist/specialists across the four sample sites, while Hurlbert's B_N was used to detect generalist/specialists with regards to temperature or pH. An adjusted p-value cut-off of 0.05 was used, and thresholds for generalist/specialist were determined from the null model distribution produced by the package during nice breadth calculations. Additionally, we calculated social niche breadth scores for ASVs using the script calculate_snb.py v0.2 with default parameters (von Meijenfeldt *et al.*, 2023).

To measure the strength of different ecological factors on shaping community assembly in the 16S rRNA gene sequence data, we used the following commands within the

microeco package with default settings: trans_nullmodel, cal_ses_betamntd, cal_rcbray, and cal_process; this approach is based on (Stegen *et al.*, 2013). To generate a distance-decay plot to look at the influence of geographic distance on the phylogenetic differences between ASVs, we calculated the geographical difference between sampling sites using the distm function (distVincentyEllipsoid option) in the geosphere v1.5-18 R package (Hijmans, 2022). To calculate biological differences in the microbial communities, we chose to use the Chao index, given the differences in the sample sizes in our data set (Chao *et al.*, 2004), implemented with the vegdist function in the vegan R package.

MAG Functional Analysis

Functional annotation of MAGs was conducted using METABOLIC-G (Zhou *et al.*, 2022), using the predicted proteins from each MAG as predicted by Prodigal (as above) with default parameters. Specifically, we used the full KOFAM database and set the cut-off value to assign the presence of a module to 0.75. We categorized the metabolic categories used by METABOLIC into either Pathway (sets of genes that take a metabolite from start to end-point) or Category (contains multiple distinct, not inter-connected pathways). We counted a Pathway as being present in a MAG if at least 50% of the genes in the pathway were present, and a Category as being present if there was at least one gene in that Category present.

Statistical Analyses

All statistical analyses were conducted in R (R Core Team, 2020), with plots produced in ggplot2 v3.3.6(Wickham, 2016) and edited for aesthetics in Inkscape (https://www.inkscape.org). Data analysis was preformed using tidyverse v1.3.1 (Wickham *et al.*, 2019). The ggpubr v0.6.0 package in R was used to add the results of a one-way ANOVA test (when comparing all four sites and when data was normally distributed), a one-way Krustal-Wallis test (when comparing all four sites for non-normally distributed data), or a

two-sided Mann-Whitney test (when comparing active vs. passive site types) to plots (Kassambara, 2023).

Results

Prokaryotic Community Diversity and Composition

To characterize the microbial communities inhabiting Antarctic geothermal sites, we extracted DNA from 21 soil samples from three geothermal areas in Victoria Land, Antarctica: Mt. Rittmann, Mt. Melbourne, and two sites on Mt. Erebus, Tramway Ridge (TR) and Western Crater (WC), for a total of four sampling sites (see Figure 1 for maps and site descriptions). Two of these four sites, Melbourne and WC, have little to no fumarolic activity (i.e., fumarolic passive) and lower maximal soil temperatures (maximum 50□). The other two sites, Rittmann and TR, have many actively steaming fumaroles (i.e., fumarolic active) and higher maximal soil temperatures (maximum 65□). Thus, we categorized WC and Melbourne as “passive” sites and TR and Rittmann as “active” sites. We used 16S rRNA gene amplicon sequencing to study the microbial communities present in these samples, generating 1,871 amplicon sequence variants (ASVs) from 926,649 reads (Table S1). We also conducted metagenome sequencing on four samples, one from each site, with almost all metagenomes reaching sequencing saturation (nonpareil C value of 1 for all except TR, which was 0.93; Figure S2B). Metagenome diversity, based on nonpareil diversity indices, were 17.8, 15.6, 15.3, and 15.2 for TR, WC, Melbourne, and Rittmann, respectively, all very low compared to other habitat microbiomes (17.5 for human-associated, 19 for lake water) (Rodriguez-R *et al.*, 2018). From these metagenomes, we binned 140 metagenome assembled genomes (MAGs): 42 from Melbourne, 15 from Rittmann, 23 from TR, and 60 from WC, all with completeness >80% (except for two that dropped to 77% during cleaning) and contamination <5% (Table S2).

Overall, we found that the microbial communities at WC and Melbourne (hereafter, passive sites) were more closely related to each other than to TR and Rittmann (hereafter, active sites) (Figure 2). This was most apparent from the 16S rRNA samples, where passive sites had significantly higher alpha diversity than the active sites (Figure 2A), clustered very closely together in a principal coordinates analysis (PCoA) (Figure 2B), and were more similar to each other based on abundance patterns both at the phylum (Figure S3A) and class level (Figure S3B). MAGs from passive sites also clustered together based on similarities in abundance profiles, separate from active sites (Figure S4). However, a pairwise PERMANOVA test using the 16S rRNA gene sequence data showed no significant difference between the microbial communities at any of the sites except between TR and WC ($p = 0.05$), which could be due to the small number of samples analysed at Melbourne and Rittmann. TR and Rittmann also clustered separately from each other in the PCoA plot, indicating distinct microbial communities at the two active sites (Figure 2B).

Based on the 16S rRNA gene sequence-based and metagenome-based community composition of the sites, we found that differences between site types were primarily attributable to greater abundances of members of *Crenarchaeota* (class *Nitrososphaeria*) at active sites, while passive sites had significantly greater abundances of phylum *Planctomycetota* (family *Gemmataceae*) and class *Acidobacteriae* within *Acidobacteriota* (Figure 2C-D, Figure S3). When looking for distinguishing taxa at each of the four sites, we found that TR was the easiest to distinguish from all other sites based on greater abundances of phylum *Bacteroidota* and the archaeal family *Nitrososphaeraceae* present (Figure 2E). Interestingly, both the Rittmann and WC metagenomes had much greater relative abundances of *Nitrososphaeria* compared to the corresponding 16S rRNA gene sequence-based community profile (85% vs. 20% for Rittmann, 25% vs. 5% for WC for metagenome vs. 16S rRNA, respectively), which could indicate primer selection against this group.

We used our complete set of non-dereplicated MAGs, categorized by the site the MAG was binned from, to examine genomic characteristics of microbes from each site. We found that MAGs binned from active sites were distinguishable by their significantly smaller average genome size (average: 3.1 Mbp for active, 4.8 Mbp for passive) and significantly higher predicted optimal growth temperature (OGT) (average: 50.8□ for active, 38.0□ for passive); MAGs from Melbourne had the largest average genome size and the lowest predicted OGT of the four sites (Figure 3A-B). We did not see significant differences in %GC, coding density, or predicted maximal growth rate between MAGs from the two site types or the four sites (Figure S5A-B).

When we explored the physicochemical parameters that could be influencing the microbial communities at these sites, we first found that the three volcanoes clustered separately from each other in the physicochemical data (Figure S6A), with elemental composition playing a large role in this separation (Figure S6B), reflecting the unique geological origin of each volcano. When we looked for correlations between physicochemical parameters and biological communities (16S rRNA gene sequence-based), although many parameters were significantly correlated (Figure S6C), only magnesium remained significant ($p = 0.033$) after Benjamini and Hochberg adjustment. Values for selected parameters are presented in Figure S6D and all physicochemical parameter values are given in Table S3.

Exploration of Cyanobacterial Diversity and Functions

We used the *Cyanobacteria* group as a case study to explore the influence of physicochemistry on the distribution and functioning of these important primary producers and to investigate the global uniqueness of these microorganisms. We binned an additional five cyanobacterial MAGs from the metagenome of Ritt3M-1, a cyanobacterial mat sample from Rittmann; this was necessary for our analysis as there were no cyanobacterial MAGs

recovered from the other Rittmann metagenome sample. In total, we retrieved 50 cyanobacterial ASVs and 9 cyanobacterial MAGs from our data set.

Overall, we found organisms from three families (*Nostocaceae*, *Leptolyngbyaceae*, and *Obscuribacteraceae*), but could only identify a few instances of clear correlations between physicochemical parameters and the distribution and functioning of these groups (Figure 4A, B). Supporting earlier findings at the entire microbial community scale, Tramway Ridge had a distinct *Cyanobacteria* community, with species from *Leptolyngbyaceae* and the genus *Fischerella* (family *Nostocaceae*) dominant and an absence of *Obscuribacteraceae*. *Leptolyngbyaceae* was positively correlated with gravimetric water content (GWC) and negatively correlated with nickel (Figure 4C). In general, *Nostocaceae* was the most widely distributed and had a positive relationship with nickel and sulfur and was negatively correlated with GWC (Figure 4C). Only MAGs from *Nostocaceae* had the capacity to fix nitrogen (Figure S7), which matches previously measured nitrogen fixation by this group at Melbourne (Broady *et al.*, 1987).We found that *Obscuribacteraceae* MAGs were non-photosynthetic, as previously observed for this group (Soo *et al.*, 2014), with pathways for ethanol fermentation and fatty acid degradation instead (Figure S7B).

To illuminate potential evolutionary relationships between the *Cyanobacteria* MAGs from this study and globally distributed representatives of these families from GTDB (GTDB species cluster representatives, 217 *Nostocaceae*, 38 *Leptolyngbyaceae*, and 55 *Obscuribacteraceae*), we used a phylogenomic approach (Figure S8). We observed that all MAGs from our study clustered with globally distributed members of these families, from the Mojave Desert to China to the continental subsurface. The short branch lengths we observed for our MAGs also confirming the cosmopolitan distribution of these families, as has been found previously at Tramway Ridge (Herbold *et al.*, 2024). These results are dependent on the completeness of the database used and should not be treated as conclusive evidence.

Functional Comparison of Sites

We analyzed the functional potential of potentially representative microbial communities at the four sites using the community metagenome sequences, post-assembly, looking at the abundances of 52 conserved marker genes that characterize various key metabolic pathways (Figure 3C). At all sites, hydrogenases involved in atmospheric H_2 scavenging (Group 1h and 1l NiFe hydrogenases) (Greening *et al.*, 2016) were present at low abundances.

Several key differences were notable between sites. First, TR was distinguishable by an enrichment in the number of genes involved in sulfur (SoxB, Sqr, and FCC) and nitrogen (NosZ, NxrA, NirS, NirK, NorB, and NarG) metabolisms (Figure 3C). These differences are expected given the relatively large concentrations of TN and sulfur at TR (Figure S6). A notable exception to the suite of nitrogen metabolising genes detected at TR is the lack of AMO (ammonia monooxygenase) encoding genes, an exception which has recently been accounted for (Herbold *et al.*, 2024). Melbourne and WC were highly similar to each other, except for the presence of methane monooxygenase genes at Melbourne and not WC (Figure 3C). Several MAGs in the genus *Methylocella* were found at Melbourne, confirming the potential for methane metabolism there, which aligns with large measured concentrations of methane at Melbourne (Geyer, 2021) (Table S2). Notably, Rittmann was distinguishable by an enrichment in marker genes for anaerobic metabolism (fumarate reductase, FrdA) and CO oxidation (CoxL), as well as much lower abundance of genes related to aerobic metabolism (cytochrome aa3 oxidase, CoxA; NADH-ubiquinone oxidoreductase, NuoF; and F-type ATP synthase, AtpA) (Figure 3C). The gas composition of Rittmann fumaroles has yet to be studied, but based on these results, is likely to be enriched in CO and depleted in oxygen.

Habitat Specialization and Community Assembly Factors

Following our taxonomic and functional analysis of these Antarctic geothermal sites, we asked whether there was a large community of shared microorganisms between these sites (habitat generalists) and, if so, whether dispersal between sites was an important factor in structuring these communities. To do this, we first classified 16S rRNA gene sequence ASVs and MAGs by the number of sites or (for ASVs) samples at which they were present, after confirming that all samples were similarly sampled using rarefaction (Figure S2). Based on this classification, we observed that most ASVs (90%) and MAGs (77%) in our study were only found at either one or two sites/samples (Figure S9A-C, F, I). However, these site/sample-specific ASVs and MAGs comprised only a small fraction of the total read count at all samples. Instead, it was the ASVs and MAGs that were found across all four sites (36 ASVs, 3% of all ASVs; 2 MAGs, 2% of all MAGs) that were generally the most abundant (>30% of all reads for both ASVs and MAGs) (Figure S9A, B, D, G, J). Melbourne and WC shared the most ASVs and MAGs (120 ASVs, 9% of all ASVs; 36 MAGs, 32% of all MAGs) (Figure S9A, B), consistent with their similar overall microbial community composition. TR shared very few ASVs with any other site (Figure S9A, B).

Next, we asked whether these common microorganisms were adapted to thrive at all of the sites (habitat generalists), or if their presence at all sites was a product of stochastic processes. We used a robust implementation of Levin's niche breadth (Finn *et al.*, 2020) with our 16S rRNA gene sequence data (reduced to 832 ASVs by eliminating ASVs only present in one sample) to identify specialists and generalists. Levin's niche breadth (B_N) is calculated by:

$$Levins' B_N = \frac{1}{R \sum_{i=1} p_i^2}$$

where R is the number of differing environments and p_i is what proportion of all instances of taxon j is found in environment i. B_N values close to 1 indicate broad habitat occupancy,

while values close to 1/*R* indicate habitat specialization. In our data set, we defined each of our four sites as different environments and used ASV read counts for the taxon proportions.

In contrast to the theoretical distribution of niche breadth scores (Figure S10A), surprisingly, we did not find any habitat generalists, with 95% of ASVs instead identified as specialists and the rest as intermediate (Figure 5A). Because two of our sample sets (TR and WC) were collected along temperature gradients, we re-ran the analysis with only high temperature samples (>35□), but again found 96% of ASVs were specialists; only one ASV (ASV_5, GAL15) was returned as a generalist (Figure S10B). This absence of habitat generalists held true both with a relaxed limit of quantification (LOQ) for including ASVs in the niche breadth analysis, as well as when we randomly subsampled within each site to bring the sample number from all sites to the same amount (data not shown). A recently developed social niche breadth metric (SNB (von Meijenfeldt *et al.*, 2023)) identifies specialists or generalists based on the compositional similarity of the communities that microbes inhabit. Using this metric, we found a distinct skew towards specialization (lower SNB score) in our data set, with a much longer tail on the specialist side both in raw SNB scores and in modified z-scores at all taxonomic ranks, supporting our findings of high habitat specialization (Figure S10).

A refinement to Levins' niche breadth adds in a measure of the resource availability across environments, called Hulbert's niche breadth (Finn *et al.*, 2020). This allows measurement of habitat specialism or generalism in reference to environmental gradients; in our case, we chose to examine temperature and pH, as these were identified as having the strongest correlations with community composition (Figure S6C). Again, we found a strong skew towards specialization for temperature and pH (95% specialists for temperature, 73% for pH) (Figure 5B, C). However, there were two ASVs identified as generalists for temperature alone (ASV 5, phylum GAL15) or both temperature and pH (ASV 42, phylum

Acidobacteriota, genus *Bryobacter*). When examining the abundance of these ASVs across temperature and pH, they clearly had a broader distribution across different temperature/pH values than highly specialized ASVs (Figure S10E). However, their distribution within sites is more variable (Figure S10E).

Next, we explored what ecological factors might be driving the lack of habitat generalists in our data set. First, we used a common method, beta nearest taxon index and Bray-Curtis-based Raup-Crick, to calculate contribution of different factors to community assembly in the 16S rRNA gene sequence data (Stegen *et al.*, 2013; Zhu *et al.*, 2022; DePoy and King, 2023). We found that, across all sites, ecological drift was the strongest factor (30.5%), with homogenous dispersal (i.e., similar community structure caused by high dispersal rates) was the second strongest factor (25.5%) (Figure 5D). On the other hand, 20% of the community assembly could be explained by dispersal limitation (i.e., dissimilar community structure caused by low dispersal rates) (Figure 5D). Environmental selection (either variable selection, i.e. dissimilar community structure caused by different environmental conditions, or homogenous selection, i.e. similar community structure caused by similar environmental conditions) was the weakest factor (11.4% variable selection, 12.9% homogenous selection) (Figure 5D). Similar results were found when comparing only active sites or only passive sites, although homogeneous environmental selection was much stronger within passive sites than active (Figure 5D). The importance of ecological drift and dispersal in structuring these communities was again apparent when we examined the distance-decay plot for the 16S rRNA gene sequence data, finding a relatively flat slope of the linear regression line (m = -0.0003), which points to strong dispersal between sites (Figure 5E). This shallow slope held true when Active and Passive site samples were considered separately as well (m = -0.0007 and -0.0003 for Active and Passive, respectively)

(Figure 5F, G). Slopes for similarly large-scale studies have shown much larger slopes (e.g., (Whitaker *et al.*, 2003; Green *et al.*, 2004; Locey *et al.*, 2020)).

Functional Imprint of Habitat Specialization

Finally, we explored the impact of high levels of habitat specialization on the microbial communities from these sites by exploring the genomic and functional profiles of MAGs based on site distribution. Generally, habitat generalists tend to be metabolic generalists (Barberán *et al.*, 2014; Chen *et al.*, 2021a); given the absence of habitat generalists in our data set, we predicted that the number of sites that a microbe is present at should have little to no correlation with genome size, number of unique protein domains (used as a proxy for metabolic breadth), or predicted metabolic pathways present. Rather, we would expect that most microorganisms would likely be metabolic specialists.

Indeed, we found very few significant differences (two-sided Wilcox test, Benjamini & Hochberg-adjusted p-value < 0.05) between different MAG site occupancy categories in genome size, number of unique pfam domains, or predicted metabolic pathways/categories present (Figure 6A-B). We did find that MAGs occupying three sites had significantly more metabolic pathways/categories than MAGs only occupying one site. Additionally, we observed significantly larger genome sizes and numbers of unique pfam domains in MAGs present at both WC and Melbourne compared with those occupying only WC (Figure 6B). Correlations between other genome characteristics and site occupancy were non-significant (Figure S11A-B). When we explored the ASVs that were identified previously as generalists for temperature and pH, ASVs 5 and 42, we were able to match ASV 5 to the MAG Ritt_bin.6.strict_final with 100% identity. This MAG had some hallmarks of generalism, with 20.9 predicted metabolic pathways/categories but a genome size of only 3.4 Mbp. Two MAGs (Erebus-WC_bin.32.strict_final and Mel_bin.5.orig_final) were classified to the *Bryobacteraceae* family, as was ASV 42; neither MAG contained any 16S sequence, so we

cannot be sure if they match ASV 42. Still, both MAGs have genome sizes and numbers of predicted metabolic pathways/categories on the upper end of the distribution (6.2 Mbp and 10.1 Mbp, respectively for genome size; 18.2 and 26.2, respectively, for metabolic pathways/categories).

We investigated the different kind of metabolisms present in MAGs based on site occupancy patterns and found few patterns (Figure 6C, Table S4). We did observe a trend towards MAGs present at two or three sites being more likely to have the capacity for nitrile hydration (6%, 17%, or 11% of MAGs present at one, two, or three sites, respectively) (Figure 6C). Nitrile hydration provides microorganisms with another avenue of acquiring carbon and nitrogen needed for growth (Fang *et al.*, 2015). Notably, the two MAGs that were present at all four sites lack many pathways/categories that were found in other MAGs, such as ethanol fermentation, nitric and nitrous oxide reduction, urea utilization, hydrogenases, and halogenated compound utilization, among others (Figure 6C), indicating that presence at all four sites does not require metabolic generalism.

As we initially hypothesized, there was no observable trend towards metabolic specialization, rather, relatively even numbers of metabolic specialists and generalists across sites and site occupancy patterns (Figure 6C). Rather, there were relatively even numbers of metabolic generalists (many types of pathways, aerobic/anaerobic, multiple energy sources, etc.) and metabolic specialists between the different site occupancy patterns. One example of a metabolic generalist, the MAG with the second largest number of metabolic pathways present, Ritt_bin.7.strict_final (*Gammaproteobacteria*, genus *Paraburkholderia*), appears to be capable of, among other things, urea utilization, thiosulfate and sulfur oxidation, nitrite reduction to ammonia, fatty acid degradation, ethanol fermentation, some types of complex carbon degradation and C1 metabolism, carbon fixation (via RuBisCo), and aerobic CO oxidation, yet was only present at Rittmann (Figure 6C). On the other hand, we observed

multiple metabolic generalist MAGs that also had broad habitat ranges, such as Erebus-WC_bin.32.strict_final (*Acidobacteriota*, order *Bryobacterales*) which was found at all sites except TR and is capable of both organic and inorganic metabolisms, and both aerobic and anaerobic metabolism (among other things, sulfur oxidation, nitrile hydration, nitrite reduction to ammonia, iron oxidation, fatty acid degradation, ethanol fermentation, some level of complex carbon degradation, chlorite reduction, and aerobic CO oxidation) (Figure 6C).

One factor that seems to be largely negating the trends towards metabolic generalism in MAGs with higher site occupancy is the presence in our dataset of a large number of archaeal MAGs with broad habitat ranges (present at 3+ sites) with few predicted metabolic pathways and very small genomes (mean archaeal genome size: 2 Mbp). For example, Erebus-TR_bin.23.permissive_final (class *Nitrososphaeria*), present at all four sites, has a limited metabolic repertoire (genes for aerobic CO oxidation, chlorite reduction, nitrite reduction, and sulfur oxidation), with mostly aerobic pathways (Figure 6C). Similarly, Ritt_bin.9.permissive_final (class *Nitrososphaeria*, order *Conexivisphaerales*), present at all sites except Melbourne, is capable of the same set of metabolic pathways minus nitrite reduction (Figure 6C).

Discussion

In this research, we have characterized the microbial communities inhabiting remote, isolated Antarctic geothermal sites and investigated the impacts of high dispersal rates on the abundance and distribution of habitat specialists and generalists in a patchy environment. We found that lower temperature, passive sites (WC, and Melbourne) share very similar microbial communities that were dominated by *Plancomycetota* and *Acidobacteriota*, while the hotter, active sites (TR, and Rittmann), each harbored distinct communities with high abundances of *Nitrososphaeria*. We also discovered an absence of habitat generalists in our

data set, with no microbes thriving at all four sites, driven by high dispersal between sites and high ecological drift. It should be noted that “habitat generalists” here refers to generalists across Antarctic geothermal sites, which is different from “globally cosmopolitan” microbes. We also note that only a single sample from each site was sequenced for metagenomics, which may not be sufficient to capture the community heterogeneity present at these sites.

Site Characterizations

Both taxonomically (Figure 2, Figure S3, Figure S4, Figure S9A, B) and functionally (Figure 3), the microbial communities at active and passive sites are distinct from each other. The higher temperatures at active sites appear to be a major barrier preventing microorganisms from passive sites colonizing active sites. This is apparent from the significantly lower predicted optimal growth temperatures (OGT) of MAGs from passive vs. active sites as well as that of MAGs from the Rittmann mat sample (32□) vs. the Rittman surface sample (60□), which was collected directly underneath the Rittmann mat sample. It is likely that the steep temperature gradient between the air (generally below -30□) and hot soil (60-65□) is a large factor in facilitating microorganisms with lower temperature optima, such as *Cyanobacteria* and *Bacteroidota*, to survive at hot spots at Rittmann and TR.[103,104] MAGs present at both active sites and at least one passive site tended to have higher predicted OGTs than all other MAGs (Figure 6B), which has been observed before for thermophiles (i.e., there is a positive correlation between maximal growth temperature and temperature growth range (Miller, Williams, *et al.*, 2009; Weltzer and Miller, 2013)).

TR and Rittmann have additional, unique physicochemical properties (nutrient availability and soil depth) that appear to influence the microbial community composition present. The extensive phototrophic mats and moss beds present at TR that are not found at the other sites(Skotnicki *et al.*, 2001) likely provide the complex carbohydrates required for microorganisms from *Bacteroidota* to thrive there, despite their low predicted OGT (35□).

The high levels of total nitrogen (TN) at TR (likely, in part, due to the extensive moss beds observed) could explain the presence of *Leptolyngbyaceae* there and its absence from other sites, given the inability of MAGs from this family in our data set to fix nitrogen. Finally, Rittmann has lower richness than other sites, both in terms of 16S rRNA gene sequence data (Figure 2A) and number of MAGs binned (Rittmann was lowest at 15), perhaps due to the shallow layer of topsoil present caused by the steep aspect. Shallow depth to bedrock has been previously associated with lower bacterial counts and diversity (Cartwright and Hui, 2015; Cartwright *et al.*, 2016). Rittmann has previously been found to have a lower diversity of *Cyanobacteria* than Melbourne and TR (Bargagli *et al.*, 1996).

It is challenging to put the functional characteristics of Antarctic geothermal microbial communities in a global context, given the scarcity of metagenomics studies examining microbial communities inhabiting geothermally heated soils (not hot spring aqueous or sediment) in other parts of the world. However, compared to globally distributed hot springs (Shu and Huang, 2021), we found Antarctic geothermal systems to be notably low in sulphur and associated metabolisms (Figure S6D, Table S3). For instance, in Yellowstone hot springs, genes associated with sulphur metabolism (sulfur and sulfate reduction, reduced sulfur oxidation, and sulfite oxidation) were found to be among the most abundant metabolic genes (Inskeep *et al.*, 2010), with similar results found in other geothermal systems globally (Poddar and Das, 2018; Power *et al.*, 2018b; Prieto-Barajas *et al.*, 2018; Wilkins *et al.*, 2019). Erebus has long been known to be a low-sulphur volcano, but the reason behind this phenomenon have yet to be untangled (Sweeney *et al.*, 2008; Sims *et al.*, 2021). On the other hand, we observed large abundances of hydrogenases and carbon monoxide dehydrogenase genes at all our sites. Other geothermal systems globally have been observed to have similarly high abundances of genes related to metabolizing geothermally produced gases, especially hydrogen (Boyd *et al.*, 2010; Lindsay *et al.*, 2019; Shu and Huang, 2021).

Interestingly, non-geothermally heated soils in Antarctica are known to be populated by a number of microbes that use trace gas (H_2, CO, and CO_2) metabolism for survival, forming the base of the food chain along with photosynthesis, resulting in similarly high abundances of these phenotypes as observed in our samples (Ji *et al.*, 2017), but with a different origin for the gases (volcanic vs. atmospheric). In fact, some of the same, broad microbial groups conducting trace gas metabolism in cold Antarctic soils are present in our samples (*Ca.* Dormibacterota and *Ca.* Eremiobacterota).

Evidence of Dispersal Leading to High Habitat Specialization

We had hypothesized that high rates of dispersal between Antarctic geothermal sites would result in a large number of shared microbes (habitat generalists) across all sites. Although we did, indeed, find evidence of high dispersal between sites (25% of all variation in our 16S rRNA gene sequence data can be explained by homogenous dispersal), we found support for the opposite of our hypothesis: there were no habitat generalists between sites, with the only shared microbial taxa present being clearly due to mass effects, not adaptation to multiple sites. This lack of habitat generalists cannot be explained by environmental selection, as variable environmental selection only accounted for 11% of community differences between all sites and there were very few environmental parameters significantly correlating with the overall biological community, and *Cyanobacteria* in particular (Figure 6C). Our results stand in stark contrast to past studies of thermophiles(Sharp *et al.*, 2014; Power *et al.*, 2018a; Louca, 2022; DePoy and King, 2023) and habitat specialists(Pandit *et al.*, 2009; Xu *et al.*, 2022) that have found environmental selection to be the primary factor influencing distribution, not dispersal, although one study did find a large influence of dispersal on the abundant biosphere of Tibetan hot springs (Zhang *et al.*, 2018). We note that many thermophilic habitat generalists have been identified before, even along strong environmental gradients, e.g. (Weltzer and Miller, 2013; Wall *et al.*, 2014; He *et al.*, 2023;

Sriaporn *et al.*, 2023), so an absence of habitat generalists in our data set is surprising, although many thermophiles tend to be habitat specialists for temperature (Miller, Strong, *et al.*, 2009; Mahato *et al.*, 2019; DePoy and King, 2023).

Our findings provide key insights into how microbial communities assemble, with high dispersal rates and ecological drift leading to a lack of habitat generalists. These results are similar to past macroecological studies that have found specialist species are favored when dispersal is high and conditions are stable (i.e., low short-term habitat disturbance) (Odum, 1969; Wilson and Yoshimura, 1994; Marvier *et al.*, 2004; Büchi and Vuilleumier, 2014; Koffel *et al.*, 2022). High dispersal rates allow specialist species to reach optimal niches and outcompete generalist species. This does not preclude the existence of globally cosmopolitan microbes at these sites, as has been postulated previously (Herbold, Lee, *et al.*, 2014), given that cosmopolitan species may be specialists for certain environments. Additionally, highly stable conditions like those experienced at these sites (Soo *et al.*, 2009), partially a result of their isolation and the absence of outside disturbances, results in few opportunities for habitat generalists to take over niches vacated by specialists during a disturbance. High environmental stability has been found in macroecology to result in a selection towards K-strategists (lower growth rate, few offspring) (Odum, 1969), which may be a contributing factor in the slow predicted growth rates of MAGs in our data set (average across all four sites: 9.2 d^{-1}, fastest: 1.9 d^{-1}). The other factor influencing the lack of habitat generalists in our data set is the large effect of ecological drift (30% of community assembly), likely due to the small population size at these sites. In macroecology, ecological drift can override niche effects in small populations, resulting in a co-occurrence of two species even when one is a superior competitor (Orrock and Watling, 2010).

The high level of habitat specialization at these sites may also be indicative of high levels of endemism, as these two have been found to be correlated in macroecological studies

(Cowling and Holmes, 1992; Ibarra and Martin, 2015). Antarctica is home to an unusually high level of endemic organisms due to its geographic isolation (Vyverman *et al.*, 2010). Tramway Ridge has previously been found to harbor a number of endemic microbes (Herbold *et al.*, 2024). Additionally, the most abundant organism across all sites was a highly divergent member of the class *Nitrososphaeria* that lacks genes for ammonia oxidation, despite this class being primarily known for its ammonia oxidation activities. The functional potential of this MAG has previously been inferred, with the proposed name *Candidatus* Australlarchaeum erebusii (Herbold *et al.*, 2024). Despite its high abundance, this MAG (Erebus-TR_bin.23.permissive_final) had the smallest genome size (1.4 mbp) in our data set. Culturing efforts are needed to better understand the mechanisms by which this novel archaeon is capable of wide habitat range despite limited metabolic flexibility, especially given that *Archaea* are generally much more dispersal limited than *Bacteria* (Louca, 2022). Our hope is that further microbial explorations at these unique sites, with broader sampling at more areas and further metagenomic, transcriptomic, and culturing work, will help uncover the wealth of novel microorganisms present.

Acknowledgments

The author(s) wish to acknowledge the contribution of NeSI to the results of this research. New Zealand's national compute and analytics services and team are supported by the New Zealand eScience Infrastructure (NeSI) and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment. URL http://www.nesi.org.nz. We especially recognize the logistical support from Antarctica New Zealand for fieldwork on Mt. Erebus and support in the field from Jon Tyler. We would like to give a big thanks to Roanna Richards-Babbage for DNA extractions and preparing samples for ICP-MS, TC/TN, and TOC/TON.

Data Availability

All R scripts used to analyze the data are available on Github at https://github.com/ThermophileResearchUnit/Antarctic_volcano_comparison_manuscript. 16S rRNA gene sequence data has been deposited at DDBJ/EMBL/GenBank under the accession KIGX00000000. The version described in this paper is the first version, KIGX01000000. Metagenome reads are available on SRA under BioProject ID PRJNA1113340.

Funding

This work was supported by the Royal Society of New Zealand (Marsden Grant 18-UOW-028 to SC, MS, IM, and CL).

References

Allan, R.N., Lebbe, L., Heyrman, J., De Vos, P., Buchanan, C.J., and Logan, N.A. (2005) Brevibacillus levickii sp. nov. and Aneurinibacillus terranovensis sp. nov., two novel thermoacidophiles isolated from geothermal soils of northern Victoria Land, Antarctica. *Int J Syst Evol Microbiol* **55**: 1039–1050.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., et al. (2014) Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.

Andrews, S. FastQC.

Archer, S.D.J., Lee, K.C., Caruso, T., Maki, T., Lee, C.K., Cary, S.C., et al. (2019) Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nat Microbiol* **4**: 925–932.

Barberán, A., Ramirez, K.S., Leff, J.W., Bradford, M.A., Wall, D.H., and Fierer, N. (2014) Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol Lett* **17**: 794–802.

Bargagli, R., Broady, P.A., and Walton, D.W.H. (1996) Preliminary investigation of the thermal biosystem of Mount Rittmann fumaroles (northern Victoria Land, Antarctica). *Antarct Sci* **8**: 121–126.

Bargagli, R., Skotnicki, M.L., Marri, L., Pepi, M., Mackenzie, A., and Agnorelli, C. (2004) New record of moss and thermophilic bacteria species and physico-chemical properties of geothermal soils on the northwest slope of Mt. Melbourne (Antarctica). *Polar Biol* **27**: 423–431.

Bell, T.H. and Bell, T. (2021) Many roads to bacterial generalism. *FEMS Microbiol Ecol* **97**: 1–10.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Boyd, E.S., Hamilton, T.L., Spear, J.R., Lavin, M., and Peters, J.W. (2010) [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism. *ISME J* **4**: 1485–1495.

Broady, P., Given, D., Greenfield, L., and Thompson, K. (1987) The biota and environment of fumaroles on Mt Melbourne, northern Victoria Land. *Polar Biol* **7**: 97–113.

Brown, C.T. and Irber, L. (2016) sourmash: a library for MinHash sketching of DNA. *The Journal of Open Source Software* **1**: 27.

Brown, J.H. (1984) On the Relationship between Abundance and Distribution of Species. *Am Nat* **124**: 255–279.

Brown, J.S. and Pavlovic, N.B. (1992) Evolution in heterogeneous environments: Effects of migration on habitat specialization. *Evol Ecol* **6**: 360–382.

Büchi, L. and Vuilleumier, S. (2014) Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. *American Naturalist* **183**: 612–624.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

Caporaso, J.G., Ackermann, G., Apprill, A., Bauer, M., Berg-Lyons, D., Betley, J., et al. (2018) Earth microbiome project: EMP 16S Illumina Amplicon Protocol. *ProtocolsIo* 1–7.

Cartwright, J., Dzantor, E.K., and Momen, B. (2016) Soil microbial community profiles and functional diversity in limestone cedar glades. *Catena (Amst)* **147**: 216–224.

Cartwright, J. and Hui, D. (2015) Soil respiration patterns and controls in limestone cedar glades. *Plant Soil* **389**: 157–169.

Chao, A., Chazdon, R.L., Colwell, R.K., and Shen, T.-J. (2004) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* **8**: 148–159.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022) GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**: 5315–5316.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Chen, Y.J., Leung, P.M., Wood, J.L., Bay, S.K., Hugenholtz, P., Kessler, A.J., et al. (2021a) Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME Journal* **15**: 2986–3004.

Chen, Y.J., Leung, P.M., Wood, J.L., Bay, S.K., Hugenholtz, P., Kessler, A.J., et al. (2021b) Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME Journal* **15**: 2986–3004.

Cowling, R.M. and Holmes, P.M. (1992) Endemism and speciation in a lowland flora from the Cape Floristic Region. *Biological Journal of the Linnean Society* **47**: 367–383.

Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A., Bik, H.M., and Eisen, J.A. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**: e243.

DePoy, A.N. and King, G.M. (2023) Distribution and diversity of anaerobic thermophiles and putative anaerobic nickel-dependent carbon monoxide-oxidizing thermophiles in mesothermal soils and sediments. *Front Microbiol* **13**: 1–14.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.

Fang, S., An, X., Liu, H., Cheng, Y., Hou, N., Feng, L., et al. (2015) Enzymatic degradation of aliphatic nitriles by Rhodococcus rhodochrous BX2, a versatile nitrile-degrading bacterium. *Bioresour Technol* **185**: 28–34.

Fernandes□Martins, M.C., Colman, D.R., and Boyd, E.S. (2023) Relationships between fluid mixing, biodiversity, and chemosynthetic primary productivity in Yellowstone hot springs. *Environ Microbiol* **25**: 1022–1040.

Fierer, N. and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**: 626–631.

Finn, D.R., Yu, J., Ilhan, Z.E., Fernandes, V.M.C., Penton, C.R., Krajmalnik-Brown, R., et al. (2020) MicroNiche: An R package for assessing microbial niche breadth and overlap from amplicon sequencing data. *FEMS Microbiol Ecol* **96**: 1–12.

Fraser, C.I., Terauds, A., Smellie, J., Convey, P., and Chown, S.L. (2014) Geothermal activity helps life survive glacial cycles. *Proc Natl Acad Sci U S A* **111**: 5634–5639.

Gao, C. (2021) ggVennDiagram: A “ggplot2” Implement of Venn Diagram.

Geyer, A. (2021) Antarctic volcanism: Active volcanism overview. *Geological Society Memoir* **55**: 55–72.

Gong, X., del Río, Á.R., Xu, L., Chen, Z., Langwig, M. V., Su, L., et al. (2022) New globally distributed bacterial phyla within the FCB superphylum. *Nat Commun* **13**:.

Green, H.C., Fisher, J.C., McLellan, S.L., Sogin, M.L., and Shanks, O.C. (2016) Identification of specialists and abundance-occupancy relationships among intestinal bacteria of Aves, Mammalia, and Actinopterygii. *Appl Environ Microbiol* **82**: 1496–1503.

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., et al. (2004) Spatial scaling of microbial eukaryote diversity. *Nature* **432**: 747–750.

Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., et al. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *ISME J* **10**: 761–777.

Greening, C., Geier, R., Wang, C., Woods, L.C., Morales, S.E., McDonald, M.J., et al. (2019) Diverse hydrogen production and consumption pathways influence methane production in ruminants. *ISME J* **13**: 2617–2632.

He, Q., Wang, S., Feng, K., Michaletz, S.T., Hou, W., Zhang, W., et al. (2023) High speciation rate of niche specialists in hot springs. *ISME J*.

Herbold, C.W., Lee, C.K., McDonald, I.R., and Cary, S.C. (2014) Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nat Commun* **5**:.

Herbold, C.W., Mcdonald, I.R., and Cary, S.C. (2014) Microbial Ecology of Geothermal Habitats in Antarctica. In *Antarctic Terrestrial Microbiology: Physical and Biological Properties of Antarctic Soils*. Cowan, D. (ed). Berlin, Heidelberg: Springer, pp. 1–328.

Herbold, C.W., Noell, S.E., Lee, C.K., Vickers, C.J., Stott, M.B., Eisen, J.A., et al. (2024) Nutritional niches of endemic, aerobic heterotrophs from an isolated Antarctic terrestrial hydrothermal refugium elucidated through metagenomics. *in review*.

Hernandez, D.J., Kiesewetter, K.N., Almeida, B.K., Revillini, D., and Afkhami, M.E. (2023) Multidimensional specialization and generalization are pervasive in soil prokaryotes. *Nat Ecol Evol*.

Hijmans, R.J. (2022) geosphere: Spherical Trigonometry.

Hughes, K.A., McCartney, H.A., Lachlan-Cope, T.A., and Pearce, D.A. (2004) A preliminary study of airborne microbial biodiversity over Peninsular Antarctica. *Cell Mol Biol (Noisy-le-grand)* **50**: 537–42.

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. *BMC Bioinformatics* **11**:.

Hyatt, D., LoCascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223–2230.

Ibarra, J.T. and Martin, K. (2015) Beyond species richness: an empirical test of top predators as surrogates for functional diversity and endemism. *Ecosphere* **6**: art142.

Ilanko, T., Fischer, T.P., Kyle, P., Curtis, A., Lee, H., and Sano, Y. (2019) Modification of fumarolic gases by the ice-covered edifice of Erebus volcano, Antarctica. *Journal of Volcanology and Geothermal Research* **381**: 119–139.

Inskeep, W.P., Rusch, D.B., Jay, Z.J., Herrgard, M.J., Kozubal, M.A., Richardson, T.H., et al. (2010) Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function. *PLoS One* **5**: e9773.

Ji, M., Greening, C., Vanwonterghem, I., Carere, C.R., Bay, S.K., Steen, J.A., et al. (2017) Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**: 400–403.

Jiang, X., Van Horn, D.J., Okie, J.G., Buelow, H.N., Schwartz, E., Colman, D.R., et al. (2022) Limits to the three domains of life: lessons from community assembly along an Antarctic salinity gradient. *Extremophiles* **26**: 15.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.

Kassambara, A. (2023) ggpubr: “ggplot2” Based Publication Ready Plots.

Koffel, T., Umemura, K., Litchman, E., and Klausmeier, C.A. (2022) A general framework for species□abundance distributions: Linking traits and dispersal to explain commonness and rarity. *Ecol Lett* **25**: 2359–2371.

Kolde, R. (2019) pheatmap: Pretty Heatmaps.

Koslicki, D., White, S., Ma, C., and Novikov, A. (2023) YACHT: an ANI-based statistical test to detect microbial presence/absence in a metagenomic sample. *bioRxiv*.

Lahti, L. and Shetty, S. (2017) Tools for microbiome analysis in R.

Lee, M.D. (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**: 4162–4164.

Li, G., Rabe, K.S., Nielsen, J., and Engqvist, M.K.M. (2019) Machine Learning Applied to Predicting Microorganism Growth Temperatures and Enzyme Catalytic Optima. *ACS Synth Biol* **8**: 1411–1420.

Liao, H., Liu, C., Ai, C., Gao, T., Yang, Q.-E., Yu, Z., et al. (2023) Mesophilic and thermophilic viruses are associated with nutrient cycling during hyperthermophilic composting. *ISME J* **17**: 916–930.

Lindsay, M.R., Colman, D.R., Amenabar, M.J., Fristad, K.E., Fecteau, K.M., Debes, R. V., et al. (2019) Probing the geological source and biological fate of hydrogen in Yellowstone hot springs. *Environ Microbiol* **21**: 3816–3830.

Liu, C., Cui, Y., Li, X., and Yao, M. (2021) microeco□: an R package for data mining in microbial community ecology. *FEMS Microbiol Ecol* **97**:.

Locey, K.J., Muscarella, M.E., Larsen, M.L., Bray, S.R., Jones, S.E., and Lennon, J.T. (2020) Dormancy dampens the microbial distance–decay relationship. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190243.

Logan, N.A., Lebbe, L., Hoste, B., Goris, J., Forsyth, G., Heyndrickx, M., et al. (2000) Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of Bacillus fumarioli sp. nov. *Int J Syst Evol Microbiol* **50**: 1741–1753.

Logares, R., Lindström, E.S., Langenheder, S., Logue, J.B., Paterson, H., Laybourn-Parry, J., et al. (2013) Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J* **7**: 937–948.

Louca, S. (2022) The rates of global bacterial and archaeal dispersal. *ISME Journal* **16**: 159–167.

Magurran, A.E. and Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.

Mahato, N.K., Sharma, A., Singh, Y., and Lal, R. (2019) Comparative metagenomic analyses of a high-altitude Himalayan geothermal spring revealed temperature-constrained habitat-specific microbial community and metabolic dynamics. *Arch Microbiol* **201**: 377–388.

Malard, L.A. and Guisan, A. (2023) Into the microbial niche. *Trends Ecol Evol* **xx**: 1–10.

Management Plan For Antarctic Specially Protected Area No . 175 (2014) **13**: 199–225.

Marvier, M., Kareiva, P., and Neubert, M.G. (2004) Habitat Destruction, Fragmentation, and Disturbance Promote Invasion by Habitat Generalists in a Multispecies Metapopulation. *Risk Analysis* **24**: 869–878.

von Meijenfeldt, F.A.B., Hogeweg, P., and Dutilh, B.E. (2023) A social niche breadth score reveals niche range strategies of generalists and specialists. *Nat Ecol Evol*.

Miller, S.R., Strong, A.L., Jones, K.L., and Ungerer, M.C. (2009) Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl Environ Microbiol* **75**: 4565–4572.

Miller, S.R., Williams, C., Strong, A.L., and Carvey, D. (2009) Ecological Specialization in a Spatially Structured Population of the Thermophilic Cyanobacterium Mastigocladus laminosus. *Appl Environ Microbiol* **75**: 729–734.

Myeong, N.R., Choe, Y.-H., Shin, S.C., Kim, J., Sul, W.J., and Kim, M. (2024) Genomic profiling of Antarctic geothermal microbiomes using long-read, Hi-C, and single-cell techniques. *Sci Data* **11**: 1023.

Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S., and Kyrpides, N.C. (2019) New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**: 505–510.

Niederberger, T.D., Bottos, E.M., Sohm, J.A., Gunderson, T., Parker, A., Coyne, K.J., et al. (2019) Rapid microbial dynamics in response to an induced wetting event in Antarctic dry valley soils. *Front Microbiol* **10**: 1–12.

Noell, S.E., Baptista, M.S., Smith, E., McDonald, I.R., Lee, C.K., Stott, M.B., et al. (2022) Unique Geothermal Chemistry Shapes Microbial Communities on Mt. Erebus, Antarctica. *Front Microbiol* **13**:.

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

Odum, E.P. (1969) The Strategy of Ecosystem Development: An understanding of ecological succession provides a basis for resolving man's conflict with nature. *Science (1979)* **164**: 262–270.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020) vegan: Community Ecology Package.

Orrock, J.L. and Watling, J.I. (2010) Local community size mediates ecological drift and competition in metacommunities. *Proceedings of the Royal Society B: Biological Sciences* **277**: 2185–2191.

Pandit, S.N., Kolasa, J., and Cottenie, K. (2009) Contrasts between habitat generalists and specialists: An empirical extension to the basic metacommunity framework. *Ecology* **90**: 2253–2262.

Parish, T.R. (1988) Surface winds over the Antarctic continent: A review. *Reviews of Geophysics* **26**: 169.

Park, M., Choe, Y.-H., and Kim, M. (2021) Soil Metagenome Sequences from a Geothermal Site at Mount Melbourne in Antarctica. *Microbiol Resour Announc* **10**: 7–8.

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.

Pearce, D.A., Hughes, K.A., Lachlan-Cope, T., Harangozo, S.A., and Jones, A.E. (2010) Biodiversity of air-borne microorganisms at Halley station, Antarctica. *Extremophiles* **14**: 145–159.

Pepi, M., Agnorelli, C., and Bargagli, R. (2005) Iron demand by thermophilic and mesophilic bacteria isolated from an antarctic geothermal soil. *BioMetals* **18**: 529–536.

Poddar, A. and Das, S.K. (2018) Microbiological studies of hot springs in India: a review. *Arch Microbiol* **200**: 1–18.

Poli, A., Esposito, E., Lama, L., Orlando, P., Nicolaus, G., de Appolonia, F., et al. (2006) Anoxybacillus amylolyticus sp. nov., a thermophilic amylase producing bacterium isolated from Mount Rittmann (Antarctica). *Syst Appl Microbiol* **29**: 300–307.

Power, J.F., Carere, C.R., Lee, C.K., Wakerley, G.L.J., Evans, D.W., Button, M., et al. (2018a) Microbial biogeography of 925 geothermal springs in New Zealand. *Nat Commun* **9**:.

Power, J.F., Carere, C.R., Lee, C.K., Wakerley, G.L.J., Evans, D.W., Button, M., et al. (2018b) Microbial biogeography of 925 geothermal springs in New Zealand. *Nat Commun* **9**:.

Power, J.F., Lowe, C.L., Carere, C.R., McDonald, I.R., Cary, S.C., and Stott, M.B. (2023) Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand. *Front Microbiol* **14**:.

Prieto-Barajas, C.M., Alcaraz, L.D., Valencia-Cantero, E., and Santoyo, G. (2018) Life in Hot Spring Microbial Mats Located in the Trans-Mexican Volcanic Belt: A 16S/18S rRNA Gene and Metagenomic Analysis. *Geomicrobiol J* **35**: 704–712.

R Core Team (2020) R: A language and environment for statistical computing.

Rodriguez-R, L.M., Gunturu, S., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (2018) Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems* **3**:.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2014) Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**: 629–635.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**: 1340–1351.

Schiro, G., Chen, Y., Blankinship, J.C., and Barberán, A. (2022) Ride the dust: linking dust dispersal and spatial distribution of microorganisms across an arid landscape. *Environ Microbiol* **24**: 4094–4107.

Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.

Schmieder, R. and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Sharp, C.E., Brady, A.L., Sharp, G.H., Grasby, S.E., Stott, M.B., and Dunfield, P.F. (2014) Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME Journal* **8**: 1166–1174.

Shu, W. and Huang, L. (2021) Microbial diversity in extreme environments. *Nat Rev Microbiol*.

Sims, K.W.W., Aster, R.C., Gaetani, G., Blichert-Toft, J., Phillips, E.H., Wallace, P.J., et al. (2021) Chapter 7.2 Mount Erebus. *Geological Society, London, Memoirs* **55**: 695–739.

Skotnicki, M.L., Selkirk, P.M., Broady, P., Adam, K.D., and Ninham, J.A. (2001) Dispersal of the moss Campylopus pyriformis on geothermal ground near the summits of Mount Erebus and Mount Melbourne, Victoria Land, Antarctica. *Antarct Sci* **13**: 280–285.

Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., et al. (2014) An Expanded Genomic Representation of the Phylum Cyanobacteria. *Genome Biol Evol* **6**: 1031–1045.

Soo, R.M., Wood, S.A., Grzymski, J.J., McDonald, I.R., and Cary, S.C. (2009) Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. *Environ Microbiol* **11**: 715–728.

Sriaporn, C., Campbell, K.A., Van Kranendonk, M.J., and Handley, K.M. (2023) Bacterial and archaeal community distributions and cosmopolitanism across physicochemically diverse hot springs. *ISME Communications* **3**: 80.

Sriswasdi, S., Yang, C.C., and Iwasaki, W. (2017) Generalist species drive microbial dispersion and evolution. *Nat Commun* **8**:.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., et al. (2013) Quantifying community assembly processes and identifying features that impose them. *ISME Journal* **7**: 2069–2079.

Sweeney, D., Kyle, P.R., and Oppenheimer, C. (2008) Sulfur dioxide emissions and degassing behavior of Erebus volcano, Antarctica. *Journal of Volcanology and Geothermal Research* **177**: 725–733.

Ul-Hasan, S., Bowers, R.M., Figueroa-Montiel, A., Licea-Navarro, A.F., Beman, J.M., Woyke, T., and Nobile, C.J. (2019) Community ecology across bacteria, archaea and microbial eukaryotes in the sediment and seawater of coastal Puerto Nuevo, Baja California. *PLoS One* **14**: e0212355.

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**: 158.

Vyverman, W., Verleyen, E., Wilmotte, A., Hodgson, D.A., Willems, A., Peeters, K., et al. (2010) Evidence for widespread endemism among Antarctic micro-organisms. *Polar Sci* **4**: 103–113.

Wall, C.A., Koniges, G.J., and Miller, S.R. (2014) Divergence with gene flow in a population of thermophilic bacteria: a potential role for spatially varying selection. *Mol Ecol* **23**: 3371–3383.

Weissman, J.L., Hou, S., and Fuhrman, J.A. (2021) Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc Natl Acad Sci U S A* **118**: 1–10.

Weltzer, M.L. and Miller, S.R. (2013) Ecological divergence of a novel group of Chloroflexus strains along a geothermal gradient. *Appl Environ Microbiol* **79**: 1353–1358.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science (1979)* **301**: 976–978.

Whiteley, A.S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., et al. (2012) Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91**: 80–88.

Wickham, H. (2016) Ggplot2: Elegant Graphics For Data Analysis, New York: Springer-Verlag.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019) Welcome to the Tidyverse. *J Open Source Softw* **4**: 1686.

Wilkins, L.G.E., Ettinger, C.L., Jospin, G., and Eisen, J.A. (2019) Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci Rep* **9**: 3059.

Wilson, D.S. and Yoshimura, J. (1994) On the Coexistence of Specialists and Generalists. *Am Nat* **144**: 692–707.

Wright, E.S. (2016) Using DECIPHER v2.0 to analyze big biological sequence data in R. *R Journal* **8**: 352–359.

Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607.

Xu, Q., Luo, G., Guo, J., Xiao, Y., Zhang, F., Guo, S., et al. (2022) Microbial generalist or specialist: Intraspecific variation and dormancy potential matter. *Mol Ecol* **31**: 161–173.

Zhang, Y., Wu, G., Jiang, H., Yang, J., She, W., Khan, I., and Li, W. (2018) Abundant and rare microbial biospheres respond differently to environmental and spatial factors in tibetan hot springs. *Front Microbiol* **9**: 1–16.

Zhang, Z., Zhang, L., Zhao, Z., Wang, H., and Ju, F. (2022) Deduplication Improves Cost-Efficiency and Yields of De novo Assembly and Binning of Shot-Gun Metagenomes in Microbiome Research. *bioRxiv* 2022.10.12.512008.

Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., et al. (2022) METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**: 1–22.

Zhu, J., Ma, Y., Huang, L., and Zhang, W. (2022) Homogeneous selection is not always important in bacterial community in the eutrophic enclosed bay. *Ecol Process* **11**: 27.

Figure Legends

Figure 1 Map and images of sampling sites. (A) Satellite image of Antarctica (top right inset, red circle indicates Victoria Land); main image shows Victoria Land and Ross Island. (B) Aerial picture of Mt. Erebus main crater. (C) Picture of the hot soil site sampled at Western Crater on Mt. Erebus. (D) Satellite image showing the summit of Mt. Erebus and the two sampling sites in this study, Tramway Ridge (-77.519444, 167.116389 DD) and Western Crater (-77.519167, 167.187778 DD). (E) Image of sampling site on Tramway Ridge. (F) Image of sampling on the side of Mt. Rittmann (-73.448445, 165.497444 DD). (G) Image of Mt. Rittmann. (H) Image of Mt. Melbourne. (I) Image of sampling at a hot soil site on Mt. Melbourne (-74.350016, 164.700139 DD). Satellite image in (A) is from the Antarctic Digital Database Map Viewer https://www.add.scar.org/, Open Source, using hillshade and bathymetry layers. Satellite image in (D) was purchased from DigitalGlobe Incorporated, Longmont CO USA (2019). All other images were taken by Ian McDonald and S. Craig Cary.

Figure 2 Diversity and composition of microbial communities across the four sites. (A) Alpha diversity of microbial communities based on 16S rRNA gene sequences across the four sites (Tramway Ridge (TR), Western Crater (WC), Melbourne, or Rittmann, n = 8, 6, 4, or 3 respectively) using the Shannon index of diversity. Paired significance test used an undirected Wilcox test. (B) Principal coordinates analysis (PCoA) of the microbial communities based on Unifrac similarities between 16S rRNA gene sequences. The percentages in brackets for each axis indicate the percentage of the variation in community composition explained along that axis. (C) Microbial community composition at the phylum level for both 16S rRNA gene sequencing samples (16S) and metagenome samples at the read level (Metagenome reads); singleM was used for determining taxonomy for the metagenome reads. For the 16S rRNA gene sequence data, the sequencing run used for

Tramway Ridge (TR) and Western Crater (WC) samples are denoted by the last digit, either 1 for the first (data from (Noell *et al.*, 2022)) or second sequencing run. For this plot, phylum names for metagenome reads were changed to match SILVA 16S rRNA gene taxonomy to allow for ease of comparison between the two methods (*Thermoproteota* to *Crenarchaeota*, *Chloroflexota* to *Chloroflexi*, CSP1-3 to GAL15, *Dormibacterota* to *Chloroflexi* (class AD3), *Eremiobacterota* to WPS-2). (D, E) Results of a differential abundance random forest analysis to identify taxa at any level that are most important for distinguishing between (D) active and passive sites, or (E) all four sites in the 16S rRNA gene sequence data. The plots on the left give the mean decrease in Gini value, which is a measure of the importance of that taxa to distinguishing the sites. The plot on the right shows the relative abundance of those taxa at the respective sites. A threshold of either 0.2 or 0.1 relative abundance was used for plotting ASVs for the active vs. passive comparison or all four sites comparison, respectively.

Figure 3 Genomic and functional characteristics of MAGs and metagenomes from each site. (A, B) Comparison of genome characteristics for MAGs recovered from different site categories (A) or sites (B). (A) n = 32 or 79 for active or passive, respectively (*p*-values: two-sided Wilcox test). (B) n = 19, 37, 42, or 13 for Tramway Ridge (TR), Western Crater (WC), Melbourne, or Rittmann, respectively. p-values given for the pairwise comparisons are for two-sided Wilcox tests. (C) Analysis of the functional potential of the microbial communities at the four site samples with metagenomes, based on the abundance of marker genes for different metabolic pathways. Abundance within community was calculated using RPKM values, which were normalized by the abundance of single copy marker genes within each metagenome. TR: Tramway Ridge. WC: Western Crater.

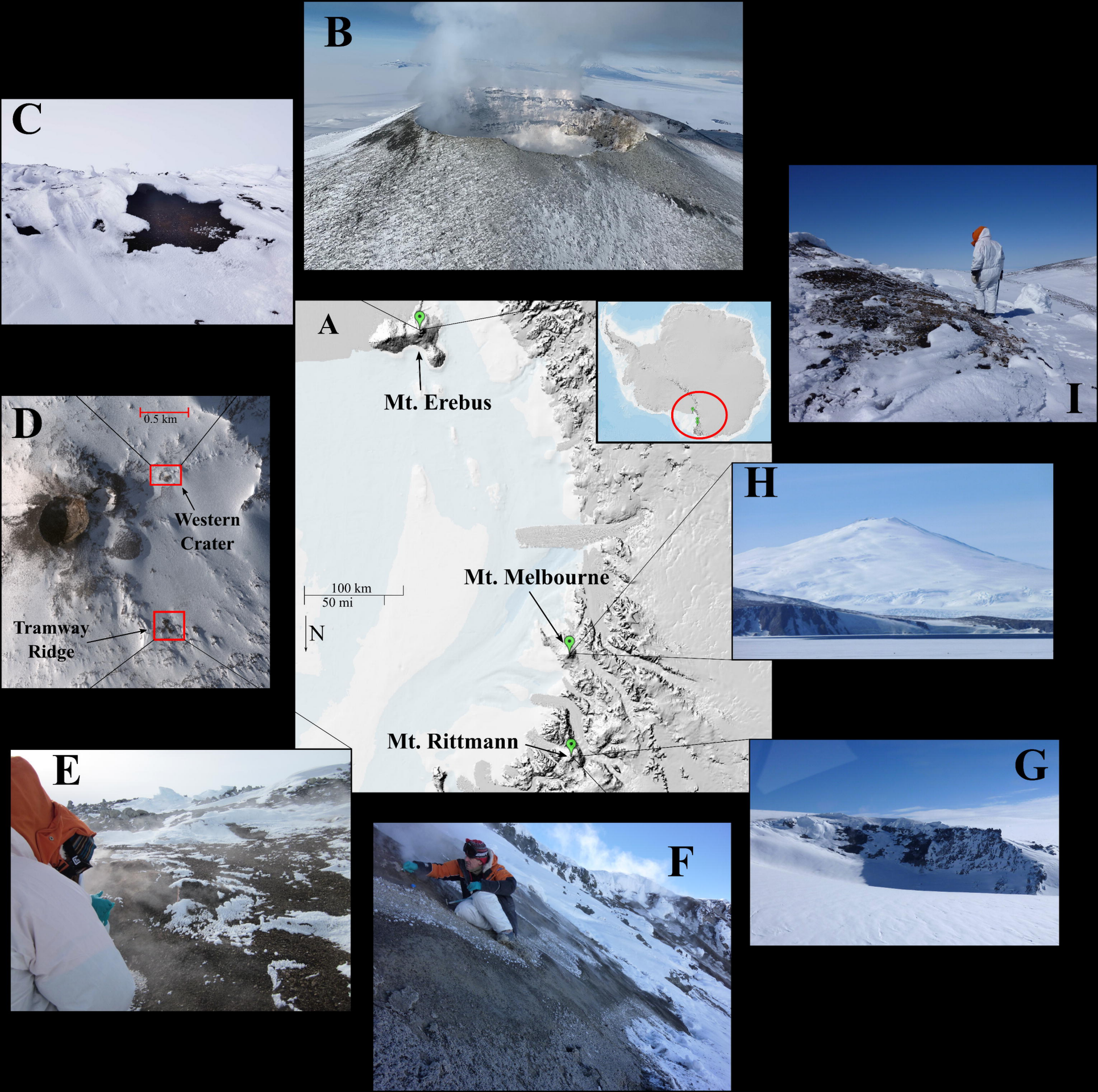
Figure 4 Analysis of Antarctic geothermal *Cyanobacteria* diversity and factors shaping it. Abundance heatmaps of *Cyanobacteria* communities using (A) 16S rRNA gene sequence and (B) metagenome assembled genome (MAG) data. For (A), relative abundances of

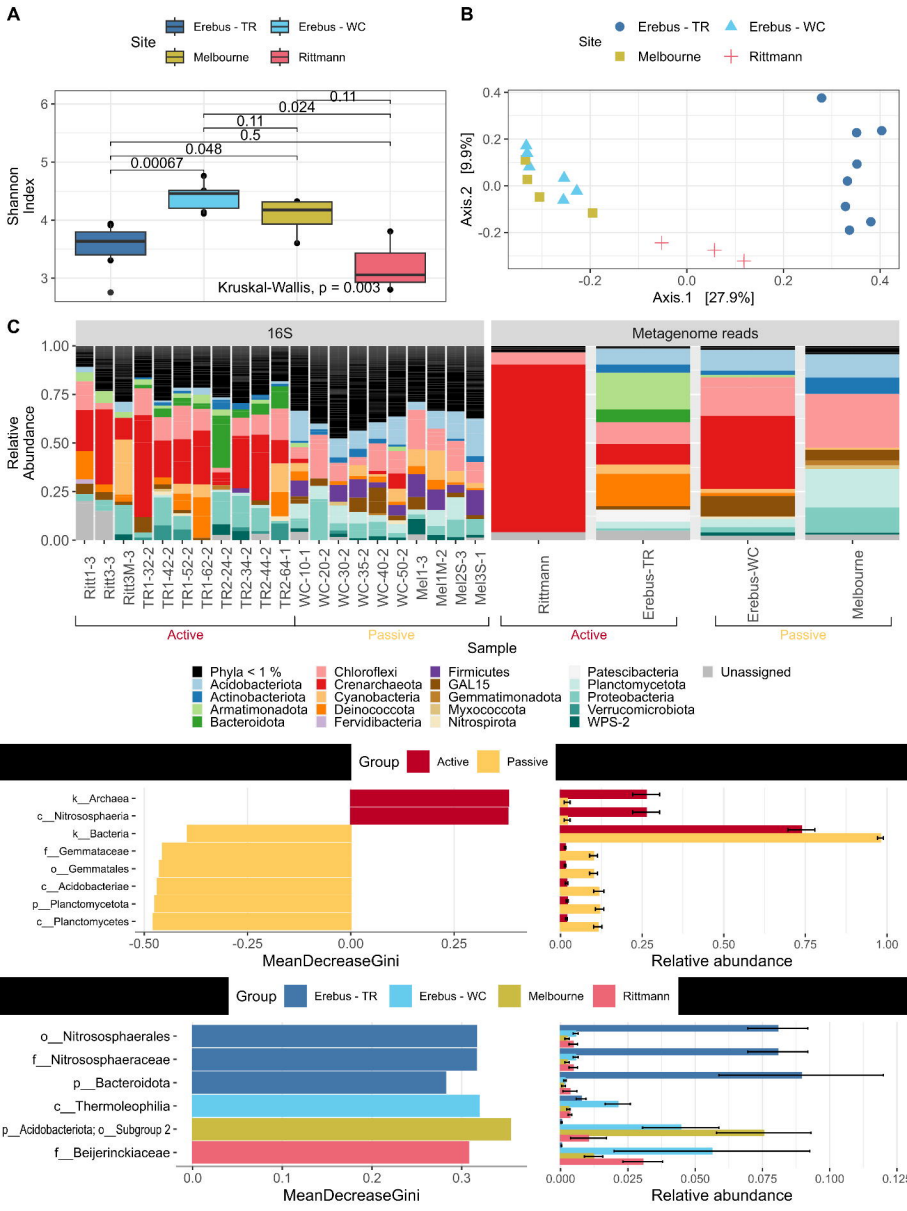
Cyanobacteria families are presented as relative to the abundance of other Cyanobacteria. For (B), MAG abundances were marked as zero if the MAG was not present at the site. Then, the abundances were normalized: $\log_{10}(\text{abundance} + 0.001) + 4$. (C) Correlation heatmap, using the Pearson coefficient, of the cyanobacterial families to the relevant environmental variables (selected for having Mantel test $p < 0.05$). GWC: gravitational water content; TN: total nitrogen.

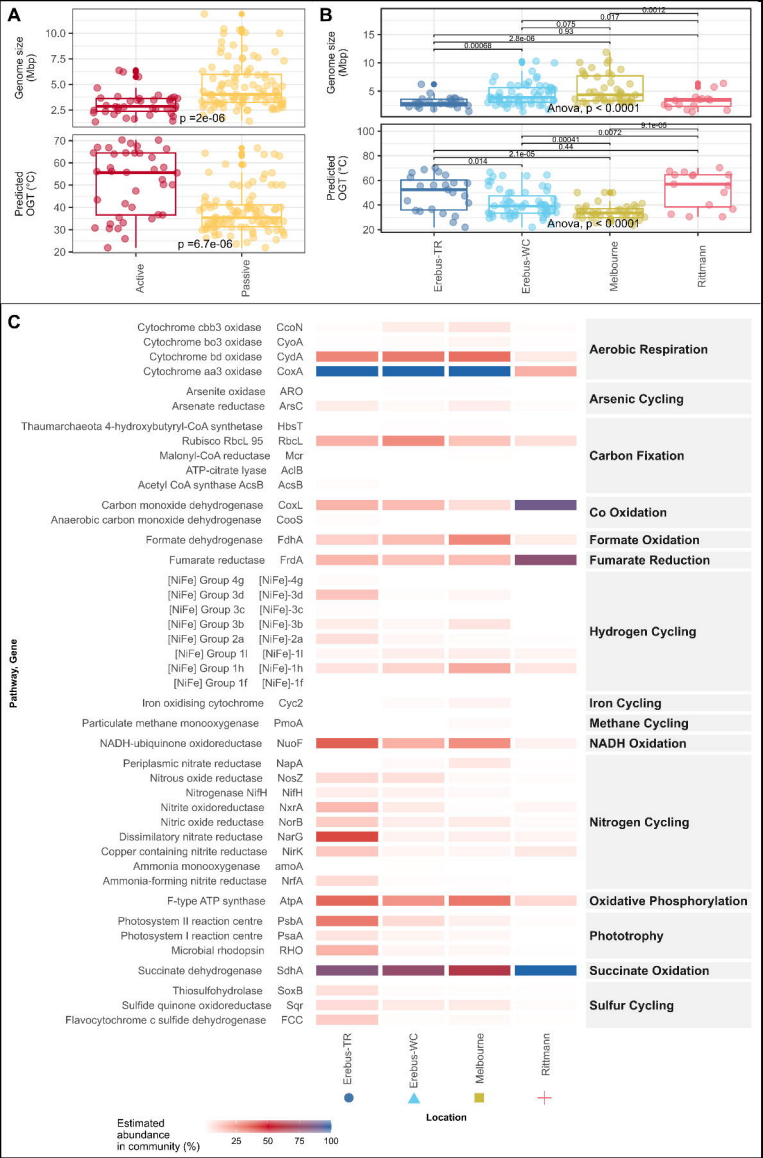
Figure 5 Ecological predictions for microorganisms at Antarctic geothermal sites. (A-C) Frequency plots identifying the number of either specialists or generalists for (A) habitat (using Levin's niche breadth, B_n), (B) temperature (using Hulbert's niche breadth, B_n), or (C) pH (using Hulbert's niche breadth, B_n) in the 16S rRNA gene amplicon sequence data. Arrows point to generalists where difficult to visualize. (D) Contribution of different ecological factors to shaping the community assembly of microorganisms in the 16S rRNA gene amplicon data. (E, F, G) Distance-decay plots for the 16S rRNA gene amplicon data with a linear regression best fit model for (E) all samples, (F) only samples from Active sites, or (G) only samples from Passive sites.

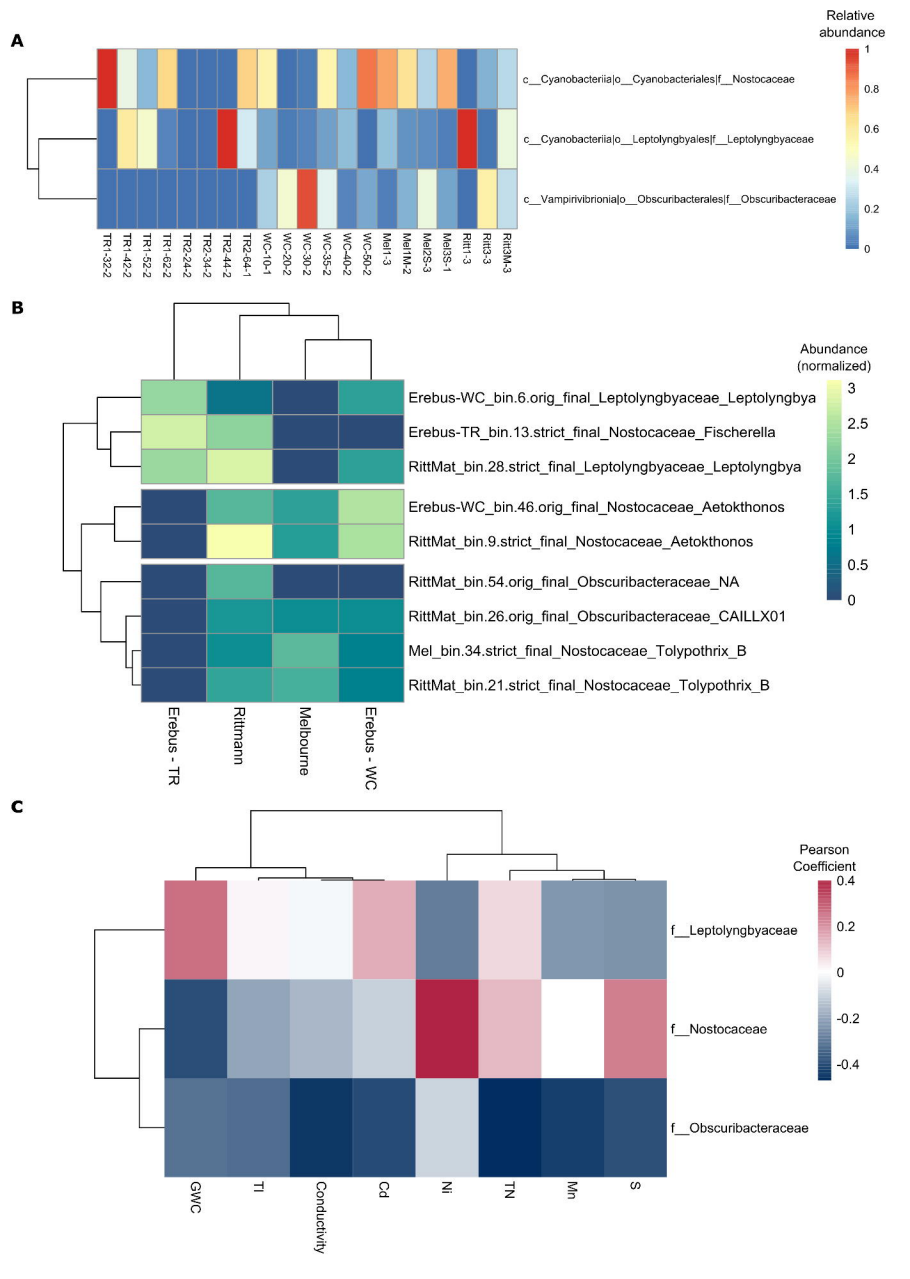
Figure 6 (A-B) Exploring evidence for genomic markers of metabolic generalism in metagenome assembled genomes (MAGs) from the four sites. For (A), n = 35, 52, 27, or 2 for the categories 1 site, 2 sites, 3 sites, or 4 sites, respectively. p-values for ANOVA tests conducted in (A) are, from top left clockwise order: 0.08, 0.07, 0.04, and 0.02. p-value shown on plot is the result of an undirect Wilcox test, Benjamini-Hochburg corrected. No differences for pfam domains remained significant after BH correction. For (B) and (C), n = 11, 2, or 7 for Tramway Ridge (TR) only, Rittmann only, or TR/Rittmann only MAGs, respectively; n = 5, 3, 8, 15, or 2 for TR/Western Crater (WC) only, WC/Rittmann only, TR/WC/Rittmann only, WC/Melbourne/Rittmann only, or MAGs at all four sites, respectively; n = 10, 11, or 36 for WC only, Melbourne only, or WC/Melbourne only MAGs,

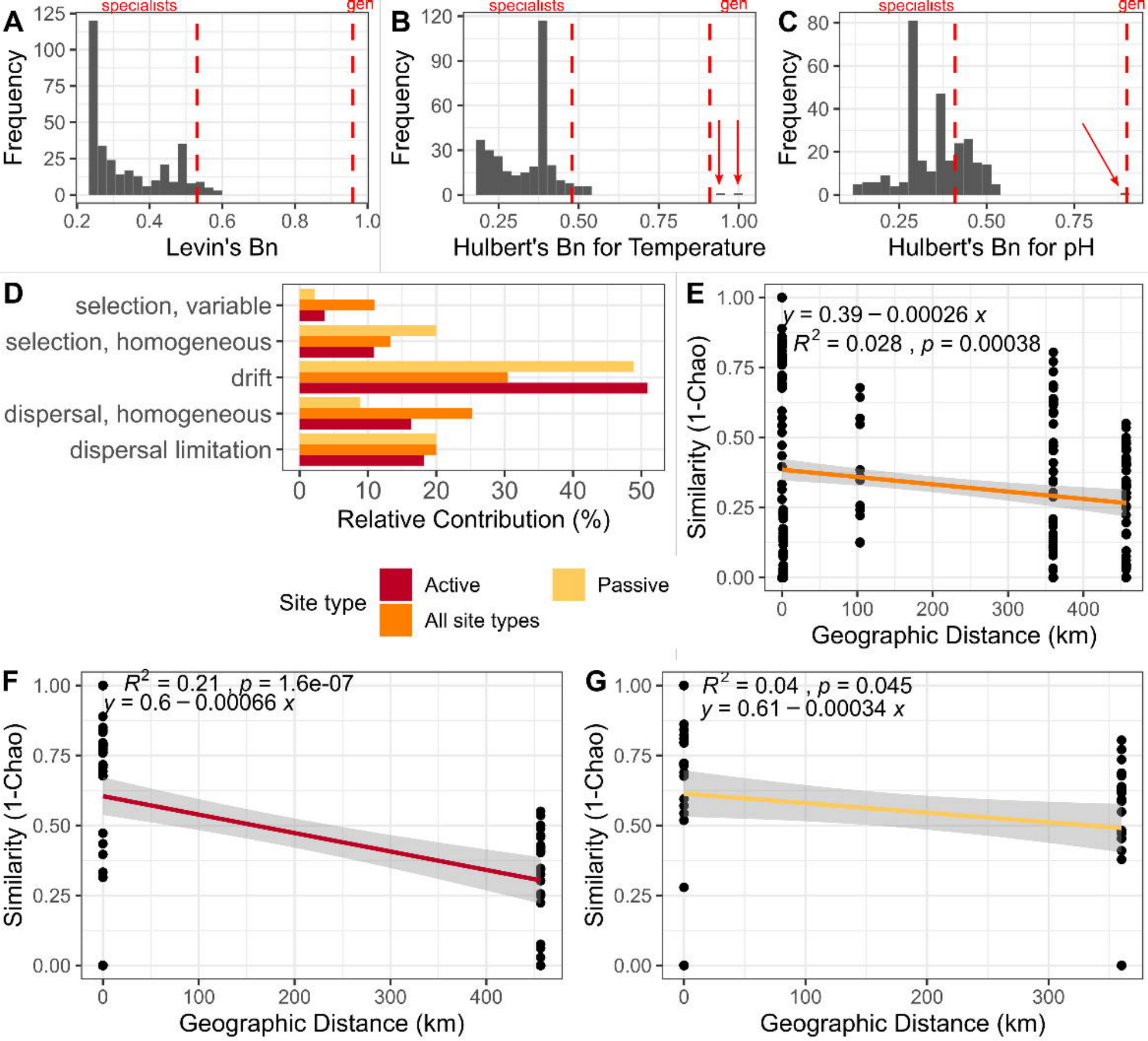
respectively. For (B), * represents a statistically significant ($p < 0.05$) result of an undirected Wilcox test after Benjamini-Hochburg correction. See Methods for how Metabolic Pathways/Categories were marked as present or absent in MAGs. (C) Heatmap showing functional predictions for MAGs based on METABOLIC predictions, separated by site occupancy (number and identity of sites occupied) and whether the metabolic predictions are categories (containing multiple, non-sequential metabolic reactions) or pathways. MAGs that are specifically mentioned in the text are indicated by an arrow. Pink boxes differentiate different site occupancy patterns. X-axis MAGs are clustered first by site occupancy pattern, then alphabetically by Phylum-Class. TR: Tramway Ridge. WC: Western Crater. RT: Rittmann. ML: Melbourne.

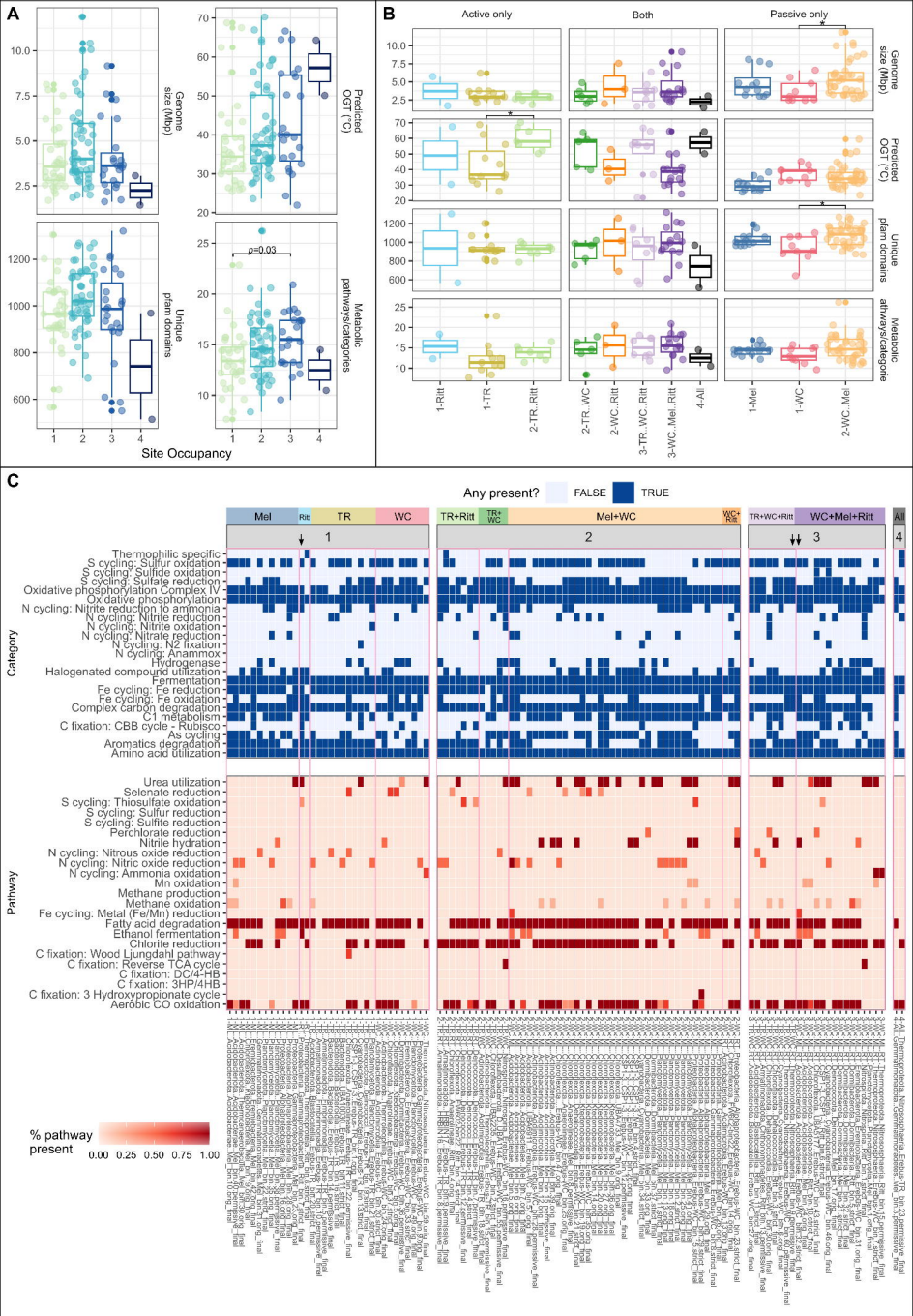

A
Mt. Erebus
Mt. Melbourne
Mt. Rittmann
100 km
50 mi
N
B
C
D
0.5 km
Western
Crater
Tramway
Ridge
E
F
G
H
I


A
Site
Erebus - TR
Erebus - WC
Melbourne
Rittmann
Shannon Index
0.11
0.024
0.11
0.5
0.048
0.00067
Kruskal-Wallis, p = 0.003
B
Site
Erebus - TR
Erebus - WC
Melbourne
Rittmann
Axis.2 [9.9%]
Axis.1 [27.9%]
C
16S
Metagenome reads
Relative Abundance
Ritt1-3
Ritt3-3
Ritt3M-3
TR1-32-2
TR1-42-2
TR1-52-2
TR1-62-2
TR2-24-2
TR2-34-2
TR2-44-2
TR2-64-1
WC-10-1
WC-20-2
WC-30-2
WC-35-2
WC-40-2
WC-50-2
Mel1-3
Mel1M-2
Mel2S-3
Mel3S-1
Rittmann
Erebus-TR
Erebus-WC
Melbourne
Active
Passive
Active
Passive
Sample
Phyla < 1 %
Acidobacteriota
Actinobacteriota
Armatimonadota
Bacteroidota
Chloroflexi
Crenarchaeota
Cyanobacteria
Deinococcota
Fervidibacteria
Firmicutes
GAL15
Gemmatimonadota
Myxococcota
Nitrospirota
Patescibacteria
Planctomycetota
Proteobacteria
Verrucomicrobiota
WPS-2
Unassigned
Group
Active
Passive
k__Archaea
c__Nitrososphaeria
k__Bacteria
f__Gemmataceae
o__Gemmatales
c__Acidobacteriae
p__Planctomycetota
c__Planctomycetes
MeanDecreaseGini
Relative abundance
Group
Erebus - TR
Erebus - WC
Melbourne
Rittmann
o__Nitrososphaerales
f__Nitrososphaeraceae
p__Bacteroidota
c__Thermoleophilia
p__Acidobacteriota; o__Subgroup 2
f__Beijerinckiaceae
MeanDecreaseGini
Relative abundance


A
Genome size (Mbp)
p =2e-06
Predicted OGT (°C)
p =6.7e-06
Active
Passive
B
Genome size (Mtap)
0.0012
0.017
0.075
0.93
2.8e-06
0.00068
Anova, p < 0.0001
Predicted OGT (°C)
9.1e-05
0.0072
0.00041
1.44
2.1e-05
0.014
Anova, p < 0.0001
Erebus-TR
Erebus-WC
Melbourne
Rittmann
C
Pathway, Gene
Cytochrome cbb3 oxidase CcoN
Cytochrome bo3 oxidase CyoA
Cytochrome bd oxidase CydA
Cytochrome aa3 oxidase CoxA
Aerobic Respiration
Arsenite oxidase ARO
Arsenate reductase ArsC
Arsenic Cycling
Thaumarchaeota 4-hydroxybutyryl-CoA synthetase HbsT
Rubisco RbcL 95 RbcL
Malonyl-CoA reductase Mcr
ATP-citrate lyase AclB
Acetyl CoA synthase AcsB AcsB
Carbon Fixation
Carbon monoxide dehydrogenase CoxL
Anaerobic carbon monoxide dehydrogenase CooS
Co Oxidation
Formate dehydrogenase FdhA
Formate Oxidation
Fumarate reductase FrdA
Fumarate Reduction
[NiFe] Group 4g [NiFe]-4g
[NiFe] Group 3d [NiFe]-3d
[NiFe] Group 3c [NiFe]-3c
[NiFe] Group 3b [NiFe]-3b
[NiFe] Group 2a [NiFe]-2a
[NiFe] Group 1l [NiFe]-1l
[NiFe] Group 1h [NiFe]-1h
[NiFe] Group 1f [NiFe]-1f
Hydrogen Cycling
Iron oxidising cytochrome Cyc2
Iron Cycling
Particulate methane monooxygenase PmoA
Methane Cycling
NADH-ubiquinone oxidoreductase NuoF
NADH Oxidation
Periplasmic nitrate reductase NapA
Nitrous oxide reductase NosZ
Nitrogenase NifH NifH
Nitrite oxidoreductase NxrA
Nitric oxide reductase NorB
Dissimilatory nitrate reductase NarG
Copper containing nitrite reductase NirK
Ammonia monooxygenase amoA
Ammonia-forming nitrite reductase NrfA
Nitrogen Cycling
F-type ATP synthase AtpA
Oxidative Phosphorylation
Photosystem II reaction centre PsbA
Photosystem I reaction centre PsaA
Microbial rhodopsin RHO
Phototrophy
Succinate dehydrogenase SdhA
Succinate Oxidation
Thiosulfohydrolase SoxB
Sulfide quinone oxidoreductase Sqr
Flavocytochrome c sulfide dehydrogenase FCC
Sulfur Cycling
Erebus-TR
Erebus-WC
Melbourne
Rittmann
Location
Estimated abundance in community (%)
25
50
75
100


A
Relative abundance
1
0.8
0.6
0.4
0.2
0
c__Cyanobacteriia|o__Cyanobacteriales|f__Nostocaceae
c__Cyanobacteriia|o__Leptolyngbyales|f__Leptolyngbyaceae
c__Vampirivibrionia|o__Obscuribacterales|f__Obscuribacteraceae
TR1-32-2
TR1-42-2
TR1-52-2
TR1-62-2
TR2-24-2
TR2-34-2
TR2-44-2
TR2-64-1
WC-10-1
WC-20-2
WC-30-2
WC-35-2
WC-40-2
WC-50-2
Mel1-3
Mel1M-2
Mel2S-3
Mel3S-1
Ritt1-3
Ritt2-3
Ritt3M-3
B
Abundance (normalized)
3
2.5
2
1.5
1
0.5
0
Erebus-WC_bin.6.orig_final_Leptolyngbyaceae_Leptolyngbya
Erebus-TR_bin.13.strict_final_Nostocaceae_Fischerella
RittMat_bin.28.strict_final_Leptolyngbyaceae_Leptolyngbya
Erebus-WC_bin.46.orig_final_Nostocaceae_Aetokthonos
RittMat_bin.9.strict_final_Nostocaceae_Aetokthonos
RittMat_bin.54.orig_final_Obscuribacteraceae_NA
RittMat_bin.26.orig_final_Obscuribacteraceae_CAILLX01
Mel_bin.34.strict_final_Nostocaceae_Tolypothrix_B
RittMat_bin.21.strict_final_Nostocaceae_Tolypothrix_B
Erebus - TR
Rittmann
Melbourne
Erebus - WC
C
Pearson Coefficient
0.4
0.2
0
-0.2
-0.4
f__Leptolyngbyaceae
f__Nostocaceae
f__Obscuribacteraceae
GWC
TI
Conductivity
Cd
Ni
TN
Mn
S


A
specialists
gen
Frequency
Levin's Bn
B
Hulbert's Bn for Temperature
C
Hulbert's Bn for pH
D
selection, variable
selection, homogeneous
drift
dispersal, homogeneous
dispersal limitation
Relative Contribution (%)
Site type
Active
All site types
Passive
E
y = 0.39 − 0.00026 x
R2 = 0.028 , p = 0.00038
Similarity (1-Chao)
Geographic Distance (km)
F
R2 = 0.21 , p = 1.6e-07
y = 0.6 − 0.00066 x
G
R2 = 0.04 , p = 0.045
y = 0.61 − 0.00034 x


A
Genome size (Mbp)
Predicted OGT (°C)
Unique pfam domains
Metabolic pathways/categories
p=0.03
Site Occupancy
B
Active only
Both
Passive only
1-Ritt
1-TR
2-TR..Ritt
2-TR..WC
2-WC..Ritt
3-TR..WC..Ritt
3-WC..Mel..Ritt
4-All
1-Mel
1-WC
2-WC..Mel
C
Any present?
FALSE
TRUE
Mel
Ritt
TR
WC
TR+Ritt
TR+WC
Mel+WC
WC+Ritt
TR+WC+Ritt
WC+Mel+Ritt
All
Category
Thermophilic specific
S cycling: Sulfur oxidation
S cycling: Sulfide oxidation
S cycling: Sulfate reduction
Oxidative phosphorylation Complex IV
Oxidative phosphorylation
N cycling: Nitrite reduction to ammonia
N cycling: Nitrite reduction
N cycling: Nitrite oxidation
N cycling: Nitrate reduction
N cycling: N2 fixation
N cycling: Anammox
Hydrogenase
Halogenated compound utilization
Fermentation
Fe cycling: Fe reduction
Fe cycling: Fe oxidation
Complex carbon degradation
C1 metabolism
C fixation: CBB cycle - Rubisco
As cycling
Aromatics degradation
Amino acid utilization
Pathway
Urea utilization
Selenate reduction
S cycling: Thiosulfate oxidation
S cycling: Sulfur reduction
S cycling: Sulfite reduction
Perchlorate reduction
Nitrile hydration
N cycling: Nitrous oxide reduction
N cycling: Nitric oxide reduction
N cycling: Ammonia oxidation
Mn oxidation
Methane production
Methane oxidation
Fe cycling: Metal (Fe/Mn) reduction
Fatty acid degradation
Ethanol fermentation
Chlorite reduction
C fixation: Wood Ljungdahl pathway
C fixation: Reverse TCA cycle
C fixation: DC/4-HB
C fixation: 3HP/4HB
C fixation: 3 Hydroxypropionate cycle
Aerobic CO oxidation
% pathway present
0.00
0.25
0.50
0.75
1.00