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 test 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 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 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.

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 (1–4)) 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 (4). Habitat generalists that have a global distribution are usually called “cosmopolitan” to distinguish them from other habitat generalists (5). 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 (1,2,6,7) and macroorganisms (8,9). 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 (10). In contrast, two reported exceptions to this trend in microbial ecology have been found under situations of severe environmental selection and dispersal limitation (e.g., strong salinity gradients (11) or the mammalian gut (12)). In microbial ecology, it is unknown which direction high dispersal rates would push the community: 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 (13,14)? 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 (15), which is likely to result in strong dispersal of microorganisms between sites within Antarctica, as has been found in the McMurdo Dry Valleys (16) and other areas of Antarctica (17,18). 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 (19). Finally, Antarctic geothermal regions serve as enrichment zones 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 (20–24). It is unknown whether

51 the strong winds in Antarctica are capable of overcoming the generally low dispersal rates for
52 thermophiles at Antarctic geothermal regions.

53 The southernmost active geothermal features in the world are located in Antarctica,
54 surrounded by vast swathes of ice and snow: Mt. Erebus, Mt. Melbourne, Mt. Rittmann, and
55 Deception Island (19). Within these geothermal areas, the interplay between cold air (as low
56 as -60°C) and hot soils results in a large variety of unique features such as ice caves, ice
57 chimneys, and ice hummocks. Of these regions, Mt. Erebus is the tallest (3794m) and most
58 volcanically active. Two of its largest geothermal sites are Tramway Ridge and Western
59 Crater, which have differing maximal temperatures (65°C vs 50°C), geothermal features
60 (many active fumaroles in bare soil producing large amounts of steam vs. warm soil mostly
61 covered by ice hummocks with less steam), physicochemistry (e.g., low pH vs. high pH), and
62 productivity levels (visible phototrophic mats and moss beds vs. primarily chemotrophic
63 driven) (25). Mt. Melbourne (2733 m tall) is located 358 km northeast of Mt. Erebus and only
64 has two geothermal fields, with maximal temperatures of 50°C, limited fumarolic activity,
65 and some moss beds (26). Geothermal activity on Mt. Rittmann (2600m tall), 450 km
66 northeast of Mt. Erebus and 103 km north of Mt. Melbourne, consists of one steep wall with
67 actively steaming fumaroles and hot soil (maximum 65°C) that is relatively shallow due to
68 the steep aspect (27). Given the strong maritime influence, presence of vascular plants, and
69 off-continent location of Deception Island north of the Antarctic circle, we excluded this site
70 from consideration in our study.

71 The microbial communities present at these unique geothermal regions have been
72 poorly incompletely characterized, especially at Mts. Rittmann and Melbourne. On Mt.
73 Erebus, culture-independent microbial community analyses have shown that cosmopolitan,
74 thermophilic microorganisms dominate the surface community of Tramway Ridge, while
75 novel and endemic members of poorly understood archaeal and bacterial groups are dominant

76 >2cm below the surface (5,28,29). Our study comparing the communities at Tramway Ridge
77 and Western Crater along thermal transects found distinct microbial communities, primarily
78 driven by strong differences in pH and water content between the two sites (25). Studies of
79 life on Mt. Melbourne have focused on macroscopic organisms (26,30,31) or were culture-
80 based (31–33), with the lone metagenomic dataset showing a community comprised of
81 *Proteobacteria*, *Chloroflexota*, and candidate phylum AD3 (now *Pseudomonadota*,
82 *Chloroflexota*, and *Dormibacteraeota* respectively) (34). Similarly, only culture-dependent
83 studies of the microorganisms at Mt. Rittmann have been conducted to-date (27,31,35,36).
84 These high-elevation, Antarctic geothermal sites have been theorized to have acted as refugia
85 for terrestrial Antarctic life during highly glaciated periods (37).

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

96 **Methods**

97 *Site Description and Sample Collection*

98 Soil samples were collected from sites on Mt. Erebus, Mt. Rittmann, and Mt.
99 Melbourne, Antarctica, between 2010 and 2019. Sampling sites with coordinates are given in
100 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 (74), with hot soil reaching 65°C in places, and has extensive moss and lichen mats covering 16% of the site (75). 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°C 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°C. 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 (30,75). 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°C and many actively steaming fumaroles with gases of unknown composition, as well as a few moss patches (75).

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 (25). 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

126 1:2.5 soil:water ratio. pH and EC measurements were made using a Thermo Scientific Orio 4-
127 Star Benchtop pH/Conductivity Meter (Thermo Fisher Scientific) using a 3-point calibration.
128 For gravimetric water content (GWC) measurements, 3 g of soil sample was dried at 105°C
129 until the sample weight remained unchanged. GWC was then calculated as the percentage of
130 weight loss during drying (5). For elemental composition (B, Na, Mg, P, K, Ca, V, Cr, Fe,
131 Mn, Co, Ni, Cu, Zn, Ga, As, Se, Sr., Ag, Cd, Ba, Tl, Pb, and U), dried samples (1g) were
132 prepared and analysed as described in (25). For total carbon and nitrogen measurements,
133 dried samples were also ground as for elemental analysis, then sent to Analytica (Hamilton,
134 New Zealand) for analysis. For total organic carbon (TOC) and nitrogen (TON)
135 measurements, samples were prepared similar to (76). Briefly, 1g of samples were weighed
136 out into acidified glass jars with cardboard disks in lids; jars and lids were washed as for the
137 ICP-MS protocol. 1 mL 6N HCl was added to samples and dried in an oven in a fume hood
138 for 24 hours at 95°C. 100 mg aliquots of samples were analysed for TOC and TON using an
139 automatic CN analyser.

140 *DNA Extraction, Amplification, and Sequencing*

141 For 16S rRNA gene amplification as well as metagenome sequencing, genomic DNA
142 was extracted from 0.5 to 0.9 g of each soil sample using a modified CTAB/bead beating
143 method (25). To have enough DNA from low biomass samples for metagenomes, multiple
144 replicates (up to 10) of samples were extracted and pooled. Genomic DNA was quantified
145 using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Massachusetts, United States).

146 The 16S rRNA gene (V4-V5 region) was amplified using PCR with the 515F (Parada)
147 and 926R (Quince) primers (77,78), adapted for Ion Torrent sequencing using fusion primers
148 with a unique tag (79). PCR reaction conditions, library prep, and 16S rRNA gene
149 sequencing are the same as described in Noell, S. E. *et al.* (2022) (25).

150 Four samples were chosen for shotgun metagenome sequencing: Ritt1-3, Ritt3M-1
151 (mat sample from Rittmann), WC-50, and Mel1-3; the mat sample was included to improve
152 the odds of recovering cyanobacterial genomes. A fifth sample (from Tramway Ridge) was
153 previously extracted and sequenced using Illumina Hi-Seq; the two surface metagenome
154 samples from this previous project (SRR6519253 and SRR6519256) were downloaded from
155 the SRA using SRAtoolkit v3.0.1 and combined into one larger sequence file. For the four
156 new metagenomic DNA samples, extracted DNA was sent to GENEWIZ Genomics Center
157 (Suzhou, China). Next generation sequencing library preparations were constructed following
158 the manufacturer's protocol (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®). For
159 each sample, 1 μg genomic DNA was randomly fragmented to <500 bp by sonication
160 (Covaris S220). The fragments were treated with End Prep Enzyme Mix for end repairing, 5'
161 Phosphorylation and dA-tailing in one reaction, followed by a T-A ligation to add adaptors to
162 both ends. Size selection of Adaptor-ligated DNA was then performed using AxyPrep Mag
163 PCR Clean-up (Axygen), and fragments of ~410 bp (with the approximate insert size of 350
164 bp) were recovered. Each sample was then amplified by PCR for 8 cycles using P5 and P7
165 primers, with both primers carrying sequences which can anneal with flowcell to perform
166 bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR
167 products were cleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated using an
168 Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by
169 Qubit2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Then libraries with different indexes
170 were multiplexed and loaded on an Illumina HiSeq instrument according to manufacturer's
171 instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150
172 paired-end (PE) configuration; image analysis and base calling were conducted by the HiSeq
173 Control Software (HCS) + RTA 2.7 (Illumina) on the HiSeq instrument.
174 *Sequence Quality Control and Taxonomic Assignment*

175 For 16S rRNA gene amplicon sequences, the raw reads from the Ion Torrent PGM
176 were processed using the DADA2 pipeline with recommended Ion Torrent parameters (80).
177 Chimeras were removed using the function removeBimeraDenovo, method "consensus."
178 Taxonomy assignment was performed using assignSpecies with the naïve Bayesian classifier
179 and SILVA database version 138.1. Eukaryotic, mitochondrial, and chloroplast sequences
180 were removed and an unrooted phylogenetic tree built using the neighbour-joining method
181 with the decipher (81) and phangorn (82) packages in R, generating final amplicon sequence
182 variants (ASVs).
183 Unfortunately, two samples (one from Tramway Ridge, TR2-64, one from Western
184 Crater, WC-10) produced no sequence data. As these two samples were sequenced in a
185 previous sequencing run, we investigated whether sequence data for these two samples from
186 the previous run could be safely used with the second sequence run samples. To control for
187 batch effects, we tested whether other samples from the two runs had similar microbial
188 communities and found that there was no pronounced batch effect for all samples between the
189 two runs (Figure S1). Thus, we decided to use sequence data for TR2-64 and WC-10 from the
190 first sequencing run alongside the second run; we denote these samples as TR2-64-1 and
191 WC-10-1 to distinguish them from samples in the second run (TR1-32-2, WC-20-2, etc.).
192 For metagenome sequences, demultiplexing was performed by bcl2fastq 2.17. Raw
193 data was filtered as follows: (1) Discard pair-end reads with adapter; (2) Discard pair-end
194 reads when the content of N bases is more than 10% in either read; (3) Discard pair-end reads
195 when the ration of bases of low quality (Q<20) is more than 0.5 in either read. All four
196 metagenome samples were quality checked using FastQC v0.11.5 (83) and the reports
197 visualized using MultiQC (84). Illumina adaptors were trimmed using trimmomatic v0.39
198 (85) and sequences were trimmed for quality using the following command:

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

201 In light of the large amount of duplicate sequences observed in some samples
202 (especially the Rittmann sample) and recent evidence that suggests deduplication is beneficial
203 (86), we used fastp v0.23.2 (87) to deduplicate all metagenome sequences using the following
204 command:
205 fastp –in1 <input file forward reads> --out1 <output forward reads> --in2 <input reverse
206 reads> --out2 <output reverse reads> -A -D -G -Q -L –html <summary file>

207 Reads with ambiguous bases were removed using prinseq-lite v0.20.4 (88) using the -
208 ns_max_n 0 command.

209 *Metagenome analysis, assembly, and binning*

210 We assessed the taxonomy of the quality trimmed metagenomes at the read level
211 using singleM v0.13.2 (https://github.com/wwood/singlem) with the singleM pipe command;
212 we saw no differences when we used the raw, non-quality trimmed sequences or the quality
213 trimmed sequences (data not shown). We used Nonpareil v3.401 (38,89) to assess the fraction
214 of community diversity sampled by the metagenome.

215 For metagenome assembled genome (MAG) assembly, binning, and processing, we
216 used the metaWRAP pipeline v1.3.2 (90). To minimize the amount of ambiguous bases in the
217 assemblies, we used the non-scaffolded contigs for binning; these non-scaffolded contigs
218 produced bins that were similar in number and quality to the scaffolded contigs (for
219 scaffolded vs non-scaffolded, the bins produced had 94.1% vs 94.0% average completeness,
220 1.5% vs 1.6% average contamination, 0.63% vs. 0.64% average GC, 41309 vs. 40696
221 average N_{50}). We used prodigal v2.6.3 in metagenome mode (91,92) to predict coding genes
222 from the non-scaffolded assemblies; these predicted coding sequences were used for the
223 metagenome functional analysis. For binning contigs in the metaWRAP pipeline, we used

224 metaBAT2 (93), MaxBin2 (94), and CONCOCT (95). For the bin_refinement module of
225 metaWRAP, we used targets of 80% and 5% for completeness and contamination,
226 respectively. Bins were then re-assembled and cleaned with MAGpurify v2.1.2 (96) using the
227 phylo-markers, clade-markers, gc-content, known-contam, and clean-bin modules to produce
228 the final MAGs. The final statistics for these MAGs were produced using checkM v1.0.12
229 (97) with the lineage_wf command. We used GTDB-Tk v2.1.1 (98) for taxonomy assignment
230 of MAGs (GTDB database R214) and both Prodigal and Prokka v1.11 (99) for predicting
231 coding gene regions. We predicted optimal growth rate of MAGs using gRodon (100) and
232 optimal growth temperature using Tome v1.0.0 (101). Finally, we used pfam_scan
233 (https://github.com/aziele/pfam_scan) to predict pfam domains in MAGs (pfam release 35.0).
234 We calculated the presence/absence of MAGs across sites by utilising MinHash
235 sketching techniques implemented with sourmash (102). Specifically, we created sourmash
236 signatures of all MAGs and quality filtered metagenomic read sets from all four sites using a
237 k-mer size of 21 and scaled to 1000. These settings allowed us to query for the presence of a
238 MAG within a site with greater sensitivity than by using a higher k-mer or scale value. We
239 selected these metrics to minimize the false negative of not finding a MAG that is present,
240 though we acknowledge this possibly increases the false positive, something recently found
241 by other groups (103). Finally, we included abundance weighting in the signature creation
242 step of the MAG and metagenome signatures so we could measure relative abundances of
243 MAGs present. To query presence/absence, we used sourmash *gather* on each MAG against
244 each metagenome to calculate containment scores. These scores are a proxy for Average
245 Nucleotide Identity where a containment score of 0.2 is equivalent to an ANI of 0.95 for a
246 signature created with a k-mer of 21. We therefore concluded that a MAG was present within
247 a metagenome if its score was greater than 0.2, to account for the varying completeness levels
248 of the MAGs and to again minimize the false negative rate. We measured MAG abundances

249 within each site using the abundance weighted percent match between the MAG and the
250 metagenome. This correlates to the proportion of metagenomic reads that would map to the
251 MAG, a proxy for abundance. The final set of MAGs was deduplicated for all analyses using
252 a 95% ANI cutoff. Where duplicate MAGs were found, the MAG from the site with its
253 highest abundance was retained.

254 *Analysis of 16S rRNA Gene Sequence Data*

255 Using 16S rRNA gene sequences, microbial community structure was assessed using
256 PcoA of center log-ratio transformed (CLR) sequence abundance on a Unifrac distance
257 matrix using the vegdist function. Statistical differences between sites were evaluated first by
258 testing samples did not differ significantly in their dispersion (using an analysis of
259 multivariate homogeneity, PERMDISP), then by using a permutational analysis of variance
260 (PERMANOVA) within the R package vegan v2.6-2 (104). Alpha diversity was also assessed
261 within vegan via the Shannon index. Venn diagrams were plotted using the R package
262 ggVennDiagram v1.2.0 (105); a custom modification to the ggVennDiagram function was
263 used to color the sections of the Venn Diagram by the abundance of ASVs in those sections.
264 Heatmaps were plotted using the R package pheatmap v1.0.12 (106). The random forest
265 differential abundance analysis was conducted using the R package microeco v0.13.0 with
266 taxa level set to "all" (107).

267 To identify physicochemical parameters that correlated with microbial community
268 structure, the mantel test function within microeco was used with default parameters. The
269 factors that had a non-adjusted two-sided p-value < 0.05 were plotted using a distance-based
270 redundancy analysis (dbRDA). To test for significant differences between physicochemical
271 parameters, the cal_diff (method = anova) and plot_diff functions within the trans_env
272 function of microeco were used. To analyse clustering of sites by physicochemical
273 parameters, principal coordinates analysis (PcoA) of physicochemical parameters in a

274 Euclidian distance matrix was performed using the vegdist function in the microbiome
275 package v.1.12.0 (108).
276 *Metagenome Functional Analysis*
277 We estimated the metabolic potential of each site using a method described by
278 Greening and colleagues (109). In short, we used DIAMOND v. 2.1.6 (query cover > 80%) to
279 search quality filtered assembled ORFs of the metagenomes against custom protein databases
280 of representative metabolic marker genes designed by the authors. The metabolic marker
281 genes include multiple pathways or energy harvesting systems, namely: sulfur cycling (AsrA,
282 FCC, Sqr, DsrA, Sor, SoxB), succinate oxidation (SdhA), selenium cycling (YgfK),
283 rhodopsin (RHO), reductive dehalogenation (RdhA), phototrophy (PsaA, PsbA, energy-
284 converting microbial rhodopsin), oxidative phosphorylation (AtpA), nitrogen cycling
285 (AmoA, HzsA, NifH, NarG, NapA, NirS, NirK, NrfA, NosZ, NxrA, NorB), NADH oxidation
286 (NuoF), methane cycling (McrA, MmoA, PmoA), isoprene oxidation (IsoA), iron cycling
287 (Cyc2, MtrB, OmcB), hydrogen cycling (catalytic subunit of [NiFe]-hydrogenases, the
288 catalytic domain of [FeFe]-hydrogenases, and Fe-hydrogenases), fumarate reduction (FrdA),
289 carbon monoxide oxidation (CoxL, CooS), carbon fixation (RbcL, AcsB, AclB, Mcr, HbsT,
290 HbsC), arsenic cycling (ARO, ArsC), and aerobic respiration (CoxA, CcoN, CyoA, CydA).
291 We filtered results based on different identity thresholds: PsaA (80%), PsbA and IsoA, AtpA,
292 ARO, YgfK (70%), HbsT (75%), group 4 [NiFe]-hydrogenases, [FeFe]-hydrogenases, CoxL,
293 AmoA, NxrA, NuoF (60%), and all others with an identity threshold of 50%. We normalized
294 the relative abundance of present markers to RPKM.
295 RPKM is recommended for relative abundance comparisons within metagenomic
296 datasets because it normalizes the data based on both sequence depth (per million reads) and
297 sequence length (in kilobases) (110). We used a three-step procedure to calculate the relative
298 abundance of each metabolic marker in the community. First, we used quality controlled

299 reads to normalize coverage of ORFs to RPKM using CoverM v0.6.1 contig
300 (https://github.com/wwood/CoverM) (with parameters: -p bwa-mem -min-read-aligned-
301 percent 0.75 -min-read-percent-identity 0.95 --min-covered-fraction 0 -m rpkm). In the
302 second step, we aligned high-quality unassembled reads to 14 universal single-copy
303 ribosomal marker genes used in SingleM v.0.12.1 (https://wwood.github.io/singlem/) and
304 PhyloSift (111) using DIAMOND (query cover > 80%, bitscore > 40) and normalized to
305 RPKM as above. Finally, we calculated the average gene copy number of each marker in the
306 community by dividing the read count to the gene (in RPKM) by the mean of the read counts
307 of the 14 universal single-copy ribosomal marker genes (in RPKM). Finally, we visualized
308 the data using ggplot2.

309 *Phylogenomic Tree Analysis*

310 For the *Cyanobacteria* phylogenomic tree, we used GtoTree v1.7.00 (112), with our 9
311 cyanobacterial MAGs along with GTDB species cluster representatives from the same three
312 Cyanobacterial families, resulting in 217 members of *Nostocaceae*, 38 for *Leptolyngbyaceae*,
313 and 55 for *Obscuribacteraceae* (accessed on 13 October 2023). We used the *Cyanobacteria*-
314 specific single copy gene (SCG) set available in the GtoTree package. The SCG threshold
315 was lowered to 35% to keep the *Obscuribacteraceae* genomes. We used four members of the
316 Terrabacteria supergroup (*Streptomyces coelicolor*, *Clostridium botulinum*, *Chloroflexus*
317 *aurantiacus*, and *Meiothermus ruber*) as the outgroups to root the tree. The tree was
318 visualized and edited using the website Interactive Tree of Life
319 (https://itol.embl.de/upload.cgi).

320 *Community Assembly Analysis*

321 To compare the abundance of cosmopolitan and site-specific microorganisms in the
322 sites/samples, ASVs were categorized by the number of sites they were present at, with the

323 threshold for presence at a site being greater than 0 relative abundance within either the
324 sample or site. MAGs were classified as present or absent as described above.
325 To calculate niche breadth for ASVs, we first eliminated ASVs only found at one site
326 (1878 ASVs reduced to 832 ASVs), equivalent to a 5% sample occupancy cutoff (stricter
327 than (13,59) but not as conservative as (113)). Next, we used the R package MicroNiche
328 v1.0.0 (42), with the default settings for the limit of quantification (LOQ) threshold (q = 1.6).
329 Levin's B_N was used to detect generalist/specialists across the four sample sites, while
330 Hurlbert's B_N was used to detect generalist/specialists with regards to temperature or pH. An
331 adjusted p-value cut-off of 0.05 was used, and thresholds for generalist/specialist were
332 determined from the null model distribution produced by the package during nice breadth
333 calculations. Additionally, we calculated social niche breadth scores for ASVs using the
334 script calculate_snb.py v0.2 with default parameters (43).
335 To measure the strength of different ecological factors on shaping community
336 assembly in the 16S rRNA gene sequence data, we used the following commands within the
337 microeco package with default settings: trans_nullmodel, cal_ses_betamntd, cal_rcbray, and
338 cal_process; this approach is based on (44). To generate a distance-decay plot to look at the
339 influence of geographic distance on the phylogenetic differences between ASVs, we
340 calculated the geographical difference between sampling sites using the distm function
341 (distVincentyEllipsoid option) in the geosphere v1.5-18 R package (114). To calculate
342 biological differences in the microbial communities, we chose to use the Chao index, given
343 the differences in the sample sizes in our data set (115), implemented with the vegdist
344 function in the vegan R package.
345 *MAG Functional Analysis*
346 Functional annotation of MAGs was conducted using METABOLIC-G (116), using
347 the predicted proteins from each MAG as predicted by Prodigal (as above) with default

348 parameters. Specifically, we used the full KOFAM database and set the cut-off value to
349 assign the presence of a module to 0.75. We categorized the metabolic categories used by
350 METABOLIC into either Pathway (sets of genes that take a metabolite from start to end-
351 point) or Category (contains multiple distinct, not inter-connected pathways). We counted a
352 Pathway as being present in a MAG if at least 50% of the genes in the pathway were present,
353 and a Category as being present if there was at least one gene in that Category present.

354 *Statistical Analyses*

355 All statistical analyses were conducted in R (117), with plots produced in ggplot2
356 v3.3.6(118) and edited for aesthetics in Inkscape (https://www.inkscape.org). Data analysis
357 was preformed using tidyverse v1.3.1 (119). The ggpubr v0.6.0 package in R was used to add
358 the results of a one-way ANOVA test (when comparing all four sites and when data was
359 normally distributed), a one-way Krustal-Wallis test (when comparing all four sites for non-
360 normally distributed data), or a two-sided Mann-Whitney test (when comparing active vs.
361 passive site types) to plots (120).

362 **Results**

363 *Prokaryotic Community Diversity and Composition*

364 To characterize the microbial communities inhabiting Antarctic geothermal sites, we
365 extracted DNA from 21 soil samples from three geothermal areas in Victoria Land,
366 Antarctica: Mt. Rittmann, Mt. Melbourne, and two sites on Mt. Erebus, Tramway Ridge (TR)
367 and Western Crater (WC), for a total of four sampling sites (see Figure 1 for maps and site
368 descriptions). Two of these four sites, Melbourne and WC, have little to no fumarolic activity
369 (i.e., fumarolic passive) and lower maximal soil temperatures (maximum 50°C). The other
370 two sites, Rittmann and TR, have many actively steaming fumaroles (i.e., fumarolic active)
371 and higher maximal soil temperatures (maximum 65°C). Thus, we categorized WC and
372 Melbourne as "passive" sites and TR and Rittmann as "active" sites. We used 16S rRNA

373 gene amplicon sequencing to study the microbial communities present in these samples,
374 generating 1,871 amplicon sequence variants (ASVs) from 926,649 reads (Table S1). We
375 also conducted metagenome sequencing on four samples, one from each site, with almost all
376 metagenomes reaching sequencing saturation (nonpareil C value of 1 for all except TR,
377 which was 0.93; Figure S2B). Metagenome diversity, based on nonpareil diversity indices,
378 were 17.8, 15.6, 15.3, and 15.2 for TR, WC, Melbourne, and Rittmann, respectively, all very
379 low compared to other habitat microbiomes (17.5 for human-associated, 19 for lake water)
380 (38). From these metagenomes, we binned 140 metagenome assembled genomes (MAGs): 42
381 from Melbourne, 15 from Rittmann, 23 from TR, and 60 from WC, all with completeness
382 >80% (except for two that dropped to 77% during cleaning) and contamination <5% (Table
383 S2).

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

397 Based on the 16S rRNA gene sequence-based and metagenome-based community
398 composition of the sites, we found that differences between site types were primarily
399 attributable to greater abundances of members of *Crenarchaeota* (class *Nitrososphaeria*) at
400 active sites, while passive sites had significantly greater abundances of phylum
401 *Planctomycetota* (family *Gemmataceae*) and class *Acidobacteriae* within *Acidobacteriota*
402 (Figure 2C-D, Figure S3). When looking for distinguishing taxa at each of the four sites, we
403 found that TR was the easiest to distinguish from all other sites based on greater abundances
404 of phylum *Bacteroidota* and the archaeal family *Nitrososphaeraceae* present (Figure 2E).
405 Interestingly, both the Rittmann and WC metagenomes had much greater relative abundances
406 of *Nitrososphaeria* compared to the corresponding 16S rRNA gene sequence-based
407 community profile (85% vs. 20% for Rittmann, 25% vs. 5% for WC for metagenome vs. 16S
408 rRNA, respectively), which could indicate primer selection against this group.
409 We used our complete set of non-dereplicated MAGs, categorized by the site the
410 MAG was binned from, to examine genomic characteristics of microbes from each site. We
411 found that MAGs binned from active sites were distinguishable by their significantly smaller
412 average genome size (average: 3.1 Mbp for active, 4.8 Mbp for passive) and significantly
413 higher predicted optimal growth temperature (OGT) (average: 50.8°C for active, 38.0°C for
414 passive); MAGs from Melbourne had the largest average genome size and the lowest
415 predicted OGT of the four sites (Figure 3A-B). We did not see significant differences in
416 %GC, coding density, or predicted maximal growth rate between MAGs from the two site
417 types or the four sites (Figure S5A-B).
418 When we explored the physicochemical parameters that could be influencing the
419 microbial communities at these sites, we first found that the three volcanoes clustered
420 separately from each other in the physicochemical data (Figure S6A), with elemental
421 composition playing a large role in this separation (Figure S6B), reflecting the unique

422 geological origin of each volcano. When we looked for correlations between physicochemical
423 parameters and biological communities (16S rRNA gene sequence-based), although many
424 parameters were significantly correlated (Figure S6C), only magnesium remained significant
425 (p = 0.033) after Benjamini and Hochberg adjustment. Values for selected parameters are
426 presented in Figure S6D and all physicochemical parameter values are given in Table S3.
427 *Exploration of Cyanobacterial Diversity and Functions*
428 We used the *Cyanobacteria* group as a case study to explore the influence of
429 physicochemistry on the distribution and functioning of these important primary producers
430 and to investigate the global uniqueness of these microorganisms. We binned an additional
431 five cyanobacterial MAGs from the metagenome of Ritt3M-1, a cyanobacterial mat sample
432 from Rittmann; this was necessary for our analysis as there were no cyanobacterial MAGs
433 recovered from the other Rittmann metagenome sample. In total, we retrieved 50
434 cyanobacterial ASVs and 9 cyanobacterial MAGs from our data set.
435 Overall, we found organisms from three families (*Nostocaceae*, *Leptolyngbyaceae*,
436 and *Obscuribacteraceae*), but could only identify a few instances of clear correlations
437 between physicochemical parameters and the distribution and functioning of these groups
438 (Figure 4A, B). Supporting earlier findings at the entire microbial community scale,
439 Tramway Ridge had a distinct *Cyanobacteria* community, with species from
440 *Leptolyngbyaceae* and the genus *Fischerella* (family *Nostocaceae*) dominant and an absence
441 of *Obscuribacteraceae*. *Leptolyngbyaceae* was positively correlated with gravimetric water
442 content (GWC) and negatively correlated with nickel (Figure 4C). In general, *Nostocaceae*
443 was the most widely distributed and had a positive relationship with nickel and sulfur and
444 was negatively correlated with GWC (Figure 4C). Only MAGs from *Nostocaceae* had the
445 capacity to fix nitrogen (Figure S7), which matches previously measured nitrogen fixation by
446 this group at Melbourne (26).We found that *Obscuribacteraceae* MAGs were non-

447 photosynthetic, as previously observed for this group (39), with pathways for ethanol
448 fermentation and fatty acid degradation instead (Figure S7B).
449 To illuminate potential evolutionary relationships between the *Cyanobacteria* MAGs
450 from this study and globally distributed representatives of these families from GTDB (GTDB
451 species cluster representatives, 217 *Nostocaceae*, 38 *Leptolyngbyaceae*, and 55
452 *Obscuribacteraceae*), we used a phylogenomic approach (Figure S8). We observed that all
453 MAGs from our study clustered with globally distributed members of these families, from the
454 Mojave Desert to China to the continental subsurface. The short branch lengths we observed
455 for our MAGs also confirming the cosmopolitan distribution of these families, as has been
456 found previously at Tramway Ridge (29). These results are dependent on the completeness of
457 the database used and should not be treated as conclusive evidence.
458 *Functional Comparison of Sites*
459 We analyzed the functional potential of the microbial communities at the four sites
460 using the community metagenome sequences, post-assembly, looking at the abundances of 52
461 conserved marker genes that characterize various key metabolic pathways (Figure 3C). All
462 sites had a high abundance of marker genes for aerobic organotrophic respiration (NuoF,
463 SdhA, AtpA, and CoxA), as well as low abundances of hydrogenases involved in
464 atmospheric H_2 scavenging (Group 1h and 1l NiFe hydrogenases) (40).
465 Several key differences were notable between sites. First, TR was distinguishable by
466 an enrichment in the number of genes involved in sulfur (SoxB, Sqr, and FCC) and nitrogen
467 (NosZ, NxrA, NirS, NirK, NorB, and NarG) metabolisms (Figure 3C; see Figure S5D for the
468 same heatmap in log scale for distinguishing lower abundance gene categories). These
469 differences are expected given the relatively large concentrations of TN and sulfur at TR
470 (Figure S6). A notable exception to the suite of nitrogen metabolising genes detected at TR is
471 the lack of AMO (ammonia monooxygenase) encoding genes, an exception which has

472 recently been accounted for (29). Melbourne and WC were highly similar to each other,
473 except for the presence of methane monooxygenase genes at Melbourne and not WC (Figure
474 3C). Several MAGs in the genus *Methylocella* were found at Melbourne, confirming the
475 potential for methane metabolism there, which aligns with large measured concentrations of
476 methane at Melbourne (41) (Table S2). Genes for methane monooxygenase genes were also
477 detected in the samples from Rittmann, but no *Methylocella* MAGs. Rittmann was
478 distinguishable by an enrichment in marker genes for anaerobic metabolism (FrdA) and CO
479 oxidation (CoxL) (Figure 3C). The gas composition of Rittmann fumaroles has yet to be
480 studied, but based on these results, is likely to be enriched in CO and methane.

481 *Habitat Specialization and Community Assembly Factors*

482 Following our taxonomic and functional analysis of these Antarctic geothermal sites,
483 we asked whether there was a large community of shared microorganisms between these sites
484 (habitat generalists) and, if so, whether dispersal between sites was an important factor in
485 structuring these communities. To do this, we first classified 16S rRNA gene sequence ASVs
486 and MAGs by the number of sites or (for ASVs) samples at which they were present, after
487 confirming that all samples were similarly sampled using rarefaction (Figure S2). Based on
488 this classification, we observed that most ASVs (90%) and MAGs (77%) in our study were
489 only found at either one or two sites/samples (Figure S9A-C, F, I). However, these
490 site/sample-specific ASVs and MAGs comprised only a small fraction of the total read count
491 at all samples. Instead, it was the ASVs and MAGs that were found across all four sites (36
492 ASVs, 3% of all ASVs; 2 MAGs, 2% of all MAGs) that were generally the most abundant
493 (>30% of all reads for both ASVs and MAGs) (Figure S9A, B, D, G, J). Melbourne and WC
494 shared the most ASVs and MAGs (120 ASVs, 9% of all ASVs; 36 MAGs, 32% of all MAGs)
495 (Figure S9A, B), consistent with their similar overall microbial community composition. TR
496 shared very few ASVs with any other site (Figure S9A, B).

497 Next, we asked whether these common microorganisms were adapted to thrive at all
498 of the sites (habitat generalists), or if their presence at all sites was a product of stochastic
499 processes. We used a robust implementation of Levin's niche breadth (42) with our 16S
500 rRNA gene sequence data (reduced to 832 ASVs by eliminating ASVs only present in one
501 sample) to identify specialists and generalists. Levin's niche breadth (B_N) is calculated by:

502 $$Levins' B_N = \frac{1}{R \sum_{i=1} p_i^2}$$

503 where R is the number of differing environments and p_i is what proportion of all instances of
504 taxon j is found in environment i. B_N values close to 1 indicate broad habitat occupancy,
505 while values close to $1/R$ indicate habitat specialization. In our data set, we defined each of
506 our four sites as different environments and used ASV read counts for the taxon proportions.
507 In contrast to the theoretical distribution of niche breadth scores (Figure S10A),
508 surprisingly, we did not find any habitat generalists, with 95% of ASVs instead identified as
509 specialists and the rest as intermediate (Figure 5A). Because two of our sample sets (TR and
510 WC) were collected along temperature gradients, we re-ran the analysis with only high
511 temperature samples (>35°C), but again found 96% of ASVs were specialists; only one ASV
512 (ASV_5, GAL15) was returned as a generalist (Figure S10B). This absence of habitat
513 generalists held true both with a relaxed limit of quantification (LOQ) for including ASVs in
514 the niche breadth analysis, as well as when we randomly subsampled within each site to bring
515 the sample number from all sites to the same amount (data not shown). A recently developed
516 social niche breadth metric (SNB(43)) identifies specialists or generalists based on the
517 compositional similarity of the communities that microbes inhabit. Using this metric, we
518 found a distinct skew towards specialization (lower SNB score) in our data set, with a much
519 longer tail on the specialist side both in raw SNB scores and in modified z-scores at all
520 taxonomic ranks, supporting our findings of high habitat specialization (Figure S10).

521 A refinement to Levins' niche breadth adds in a measure of the resource availability
522 across environments, called Hulbert's niche breadth (42). This allows measurement of habitat
523 specialism or generalism in reference to environmental gradients; in our case, we chose to
524 examine temperature and pH, as these were identified as having the strongest correlations
525 with community composition (Figure S6C). Again, we found a strong skew towards
526 specialization for temperature and pH (95% specialists for temperature, 73% for pH) (Figure
527 5B, C). However, there were two ASVs identified as generalists for temperature alone (ASV
528 5, phylum GAL15) or both temperature and pH (ASV 42, phylum *Acidobacteriota*, genus
529 *Bryobacter*). When examining the abundance of these ASVs across temperature and pH, they
530 clearly had a broader distribution across different temperature/pH values than highly
531 specialized ASVs (Figure S10E). However, their distribution within sites is more variable
532 (Figure S10E).
533 Next, we explored what ecological factors might be driving the lack of habitat
534 generalists in our data set. First, we used a common method, beta nearest taxon index and
535 Bray-Curtis-based Raup-Crick, to calculate contribution of different factors to community
536 assembly in the 16S rRNA gene sequence data (44–46). We found that, across all sites,
537 ecological drift was the strongest factor (30.5%), with homogenous dispersal (i.e., similar
538 community structure caused by high dispersal rates) was the second strongest factor (25.5%)
539 (Figure 5D). On the other hand, 20% of the community assembly could be explained by
540 dispersal limitation (i.e., dissimilar community structure caused by low dispersal rates)
541 (Figure 5D). Environmental selection (either variable selection, i.e. dissimilar community
542 structure caused by different environmental conditions, or homogenous selection, i.e. similar
543 community structure caused by similar environmental conditions) was the weakest factor
544 (11.4% variable selection, 12.9% homogenous selection) (Figure 5D). Similar results were
545 found when comparing only active sites or only passive sites, although homogeneous

546 environmental selection was much stronger within passive sites than active (Figure 5D). The
547 importance of ecological drift and dispersal in structuring these communities was again
548 apparent when we examined the distance-decay plot for the 16S rRNA gene sequence data,
549 finding a relatively flat slope of the linear regression line (m = -0.0003), which points to
550 strong dispersal between sites (Figure 5E). This shallow slope held true when Active and
551 Passive site samples were considered separately as well (m = -0.0007 and -0.0003 for Active
552 and Passive, respectively) (Figure 5F, G). Slopes for similarly large-scale studies have shown
553 much larger slopes (e.g., (47–49)).

554 *Functional Imprint of Habitat Specialization*

555 Finally, we explored the impact of high levels of habitat specialization on the
556 microbial communities from these sites by exploring the genomic and functional profiles of
557 MAGs based on site distribution. Generally, habitat generalists tend to be metabolic
558 generalists (1,2); given the absence of habitat generalists in our data set, we predicted that the
559 number of sites that a microbe is present at should have little to no correlation with genome
560 size, number of unique protein domains (used as a proxy for metabolic breadth), or predicted
561 metabolic pathways present. Rather, we would expect that most microorganisms would likely
562 be metabolic specialists.

563 Indeed, we found very few significant differences (two-sided Wilcox test, Benjamini
564 & Hochberg-adjusted p-value < 0.05) between different MAG site occupancy categories in
565 genome size, number of unique pfam domains, or predicted metabolic pathways/categories
566 present (Figure 6A-B). We did find that MAGs occupying three sites had significantly more
567 metabolic pathways/categories than MAGs only occupying one site. Additionally, we
568 observed significantly larger genome sizes and numbers of unique pfam domains in MAGs
569 present at both WC and Melbourne compared with those occupying only WC (Figure 6B).
570 Correlations between other genome characteristics and site occupancy were non-significant

571 (Figure S11A-B). When we explored the ASVs that were identified previously as generalists
572 for temperature and pH, ASVs 5 and 42, we were able to match ASV 5 to the MAG
573 Ritt_bin.6.strict_final with 100% identity. This MAG had some hallmarks of generalism,
574 with 20.9 predicted metabolic pathways/categories but a genome size of only 3.4 Mbp. Two
575 MAGs (Erebus-WC_bin.32.strict_final and Mel_bin.5.orig_final) were classified to the
576 *Bryobacteraceae* family, as was ASV 42; neither MAG contained any 16S sequence, so we
577 cannot be sure if they match ASV 42. Still, both MAGs have genome sizes and numbers of
578 predicted metabolic pathways/categories on the upper end of the distribution (6.2 Mbp and
579 10.1 Mbp, respectively for genome size; 18.2 and 26.2, respectively, for metabolic
580 pathways/categories).

581 We investigated the different kind of metabolisms present in MAGs based on site
582 occupancy patterns and found few patterns (Figure 6C, Table S4). We did observe a trend
583 towards MAGs present at two or three sites being more likely to have the capacity for nitrile
584 hydration (6%, 17%, or 11% of MAGs present at one, two, or three sites, respectively) and
585 nitrate reduction (6%, 13%, or 11% of MAGs present at one, two, or three sites, respectively)
586 (Figure 6C). Nitrile hydration provides microorganisms with another avenue of acquiring
587 carbon and nitrogen needed for growth (50). Also, MAGs at one or two sites were more
588 likely to have the capacity for selenate reduction compared to those at three or four sites (9%,
589 12%, or 0% of MAGs present at one, two, or three and four sites, respectively) (Figure 6C).
590 Notably, the two MAGs that were present at all four sites lack many pathways/categories that
591 were found in other MAGs, such as ethanol fermentation, nitric and nitrous oxide reduction,
592 urea utilization, hydrogenases, and halogenated compound utilization, among others (Figure
593 6C), indicating that presence at all four sites does not require metabolic generalism.

594 As we initially hypothesized, there was no observable trend towards metabolic
595 specialization, rather, relatively even numbers of metabolic specialists and generalists across

596 sites and site occupancy patterns (Figure 6C). Rather, there were relatively even numbers of
597 metabolic generalists (many types of pathways, aerobic/anaerobic, multiple energy sources,
598 etc.) and metabolic specialists between the different site occupancy patterns. One example of
599 a metabolic generalist, the MAG with the second largest number of metabolic pathways
600 present, Ritt_bin.7.strict_final (*Gammaproteobacteria*, genus *Paraburkholderia*), appears to
601 be capable of, among other things, urea utilization, thiosulfate and sulfur oxidation, nitrite
602 reduction to ammonia, fatty acid degradation, ethanol fermentation, some types of complex
603 carbon degradation and C1 metabolism, carbon fixation (via RuBisCo), and aerobic CO
604 oxidation, yet was only present at Rittmann (Figure 6C). On the other hand, we observed
605 multiple metabolic generalist MAGs that also had broad habitat ranges, such as Erebus-
606 WC_bin.32.strict_final (*Acidobacteriota*, order *Bryobacterales*) which was found at all sites
607 except TR and is capable of both organic and inorganic metabolisms, and both aerobic and
608 anaerobic metabolism (among other things, sulfur oxidation, nitrile hydration, nitrite
609 reduction to ammonia, iron oxidation, fatty acid degradation, ethanol fermentation, some
610 level of complex carbon degradation, chlorite reduction, and aerobic CO oxidation) (Figure
611 6C).

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

620 sites except Melbourne, is capable of the same set of metabolic pathways minus nitrite
621 reduction (Figure 6C).
622 We did observe that most MAGs only present at active sites (e.g., Rittmann only, TR
623 only, or Rittmann + TR only) were lacking in several metabolic categories that were found in
624 MAGs present in at least one passive site (Figure S11C). These pathways include methane
625 oxidation (unique to MAGs in Melbourne, matching the metagenome functional analysis
626 results), nitrile hydration, manganese cycling, nitrate reduction, urea utilization, and ethanol
627 fermentation (Figure S11C). These metabolic pathways do not match with measured,
628 corresponding nutrient concentrations (e.g., manganese and total nitrogen levels are
629 significantly greater at TR than WC or Melbourne (Figure S6D), despite manganese cycling
630 being enriched in MAGs present at WC and Melbourne compared to TR), so it is not clear
631 why these pathways are enriched in these MAGs.

632 **Discussion**

633 In this research, we have characterized the microbial communities inhabiting remote,
634 isolated Antarctic geothermal sites and investigated the impacts of high dispersal rates on the
635 abundance and distribution of habitat specialists and generalists in a patchy environment. We
636 found that lower temperature, passive sites (WC, and Melbourne) share very similar
637 microbial communities that were dominated by *Plancomycetota* and *Acidobacteriota*, while
638 the hotter, active sites (TR, and Rittmann), each harbored distinct communities with high
639 abundances of *Nitrososphaeria*. We also discovered an absence of habitat generalists in our
640 data set, with no microbes thriving at all four sites, driven by high dispersal between sites and
641 high ecological drift. It should be noted that "habitat generalists" here refers to generalists
642 across Antarctic geothermal sites, which is different from "globally cosmopolitan" microbes.
643 Both taxonomically (Figure 2, Figure S3, Figure S4, Figure S9A, B) and functionally
644 (Figure 3), the microbial communities at active and passive sites are distinct from each other.

645 The higher temperatures at active sites appear to be a major barrier preventing
646 microorganisms from passive sites colonizing active sites. This is apparent from the
647 significantly lower predicted optimal growth temperatures (OGT) of MAGs from passive vs.
648 active sites as well as that of MAGs from the Rittmann mat sample (32°C) vs. the Rittman
649 surface sample (60°C), which was collected directly underneath the Rittmann mat sample. It
650 is likely that the steep temperature gradient between the air (generally below -30°C) and hot
651 soil (60-65°C) is a large factor in facilitating microorganisms with lower temperature optima,
652 such as *Cyanobacteria* and *Bacteroidota*, to survive at hot spots at Rittmann and TR.
653 Additionally, the lack of methanotrophic activity at the active fumarolic sites is
654 understandable as thermophilic methanotrophy is only carried out by a limited number of
655 taxa at temperatures >55 °C, e.g. *Methylacidiphilum*, *Methylothermus* and putatively
656 *Bathoarchaeota* (51,52). MAGs present at both active sites and at least one passive site
657 tended to have higher predicted OGTs than all other MAGs (Figure 6B), which has been
658 observed before for thermophiles (i.e., there is a positive correlation between maximal
659 growth temperature and temperature growth range (21,53)).
660 TR and Rittmann have additional, unique physicochemical properties (nutrient
661 availability and soil depth) that appear to influence the microbial community composition
662 present. The extensive phototrophic mats and moss beds present at TR that are not found at
663 the other sites(30) likely provide the complex carbohydrates required for microorganisms
664 from *Bacteroidota* to thrive there, despite their low predicted OGT (35°C). The high levels of
665 total nitrogen (TN) at TR (likely, in part, due to the extensive moss beds observed) could
666 explain the presence of *Leptolyngbyaceae* there and its absence from other sites, given the
667 inability of MAGs from this family in our data set to fix nitrogen. Finally, Rittmann has
668 lower richness than other sites, both in terms of 16S rRNA gene sequence data (Figure 2A)
669 and number of MAGs binned (Rittmann was lowest at 15), perhaps due to the shallow layer

670 of topsoil present caused by the steep aspect. Shallow depth to bedrock has been previously
671 associated with lower bacterial counts and diversity (54,55). Rittmann has previously been
672 found to have a lower diversity of *Cyanobacteria* than Melbourne and TR (27).
673 We had hypothesized that high rates of dispersal between Antarctic geothermal sites
674 would result in a large number of shared microbes (habitat generalists) across all sites.
675 Although we did, indeed, find evidence of high dispersal between sites (25% of all variation
676 in our 16S rRNA gene sequence data can be explained by homogenous dispersal), we found
677 support for the opposite of our hypothesis: there were no habitat generalists between sites,
678 with the only shared microbial taxa present being clearly due to mass effects, not adaptation
679 to multiple sites. This lack of habitat generalists cannot be explained by environmental
680 selection, as variable environmental selection only accounted for 11% of community
681 differences between all sites and there were very few environmental parameters significantly
682 correlating with the overall biological community, and *Cyanobacteria* in particular (Figure
683 6C). Our results stand in stark contrast to past studies of thermophiles(22,45,56,57) and
684 habitat specialists(58,59) that have found environmental selection to be the primary factor
685 influencing distribution, not dispersal, although one study did find a large influence of
686 dispersal on the abundant biosphere of Tibetan hot springs (7). We note that many
687 thermophilic habitat generalists have been identified before, even along strong environmental
688 gradients, e.g. (21,60–62), so an absence of habitat generalists in our data set is surprising,
689 although many thermophiles tend to be habitat specialists for temperature (20,45,63).
690 Our findings provide key insights into how microbial communities assemble, with
691 high dispersal rates and ecological drift leading to a lack of habitat generalists. These results
692 are similar to past macroecological studies that have found specialist species are favored
693 when dispersal is high and conditions are stable (i.e., low short-term habitat disturbance) (64–
694 68). High dispersal rates allow specialist species to reach optimal niches and outcompete

695 generalist species. This does not preclude the existence of globally cosmopolitan microbes at
696 these sites, as has been postulated previously (5), given that cosmopolitan species may be
697 specialists for certain environments. Additionally, highly stable conditions like those
698 experienced at these sites (28), partially a result of their isolation and the absence of outside
699 disturbances, results in few opportunities for habitat generalists to take over niches vacated
700 by specialists during a disturbance. High environmental stability has been found in
701 macroecology to result in a selection towards K-strategists (lower growth rate, few offspring)
702 (68), which may be a contributing factor in the slow predicted growth rates of MAGs in our
703 data set (average across all four sites: 9.2 d^{-1}, fastest: 1.9 d^{-1}). The other factor influencing the
704 lack of habitat generalists in our data set is the large effect of ecological drift (30% of
705 community assembly), likely due to the small population size at these sites. In macroecology,
706 ecological drift can override niche effects in small populations, resulting in a co-occurrence
707 of two species even when one is a superior competitor (69).

708 The high level of habitat specialization at these sites may also be indicative of high
709 levels of endemism, as these two have been found to be correlated in macroecological studies
710 (70,71). Antarctica is home to an unusually high level of endemic organisms due to its
711 geographic isolation (72). Tramway Ridge has previously been found to harbor a number of
712 endemic microbes (29). Additionally, the most abundant organism across all sites was a
713 highly divergent member of the class *Nitrososphaeria* that lacks genes for ammonia
714 oxidation, despite this class being primarily known for its ammonia oxidation activities (73).
715 The functional potential of this MAG has previously been inferred, with the proposed name
716 *Candidatus* Australlarchaeum erebusii (29). Despite its high abundance, this MAG (Erebus-
717 TR_bin.23.permissive_final) had the smallest genome size (1.4 mbp) in our data set.
718 Culturing efforts are needed to better understand the mechanisms by which this novel
719 archaeon is capable of wide habitat range despite limited metabolic flexibility, especially

given that *Archaea* are generally much more dispersal limited than *Bacteria* (56). 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

1. Chen YJ, Leung PM, Wood JL, Bay SK, Hugenholtz P, Kessler AJ, et al. Metabolic

flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. ISME J [Internet]. 2021;15(10):2986–3004. Available from: http://dx.doi.org/10.1038/s41396-021-00988-w

2. Barberán A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. Ecol Lett. 2014;17(7):794–802.

3. Bell TH, Bell T. Many roads to bacterial generalism. FEMS Microbiol Ecol. 2021;97(1):1–10.

4. Malard LA, Guisan A. Into the microbial niche. Trends Ecol Evol [Internet]. 2023 May;xx(xx):1–10. Available from: https://doi.org/10.1016/j.tree.2023.04.015

5. Herbold CW, Lee CK, McDonald IR, Cary SC. Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. Nat Commun. 2014;5(May).

6. Gong X, del Río ÁR, Xu L, Chen Z, Langwig M V., Su L, et al. New globally distributed bacterial phyla within the FCB superphylum. Nat Commun. 2022;13(1).

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

8. Brown JH. On the Relationship between Abundance and Distribution of Species. Am Nat [Internet]. 1984 Aug;124(2):255–79. Available from: https://www.journals.uchicago.edu/doi/10.1086/284267

9. Magurran AE, Henderson PA. Explaining the excess of rare species in natural species abundance distributions. Nature [Internet]. 2003 Apr;422(6933):714–6. Available from: http://www.nature.com/articles/nature01547

10. Brown JS, Pavlovic NB. Evolution in heterogeneous environments: Effects of

770 migration on habitat specialization. Evol Ecol. 1992;6(5):360–82.
771 11. Logares R, Lindström ES, Langenheder S, Logue JB, Paterson H, Laybourn-Parry J, et
772 al. Biogeography of bacterial communities exposed to progressive long-term
773 environmental change. ISME J [Internet]. 2013 May 20;7(5):937–48. Available from:
774 http://www.nature.com/articles/ismej2012168
775 12. Green HC, Fisher JC, McLellan SL, Sogin ML, Shanks OC. Identification of
776 specialists and abundance-occupancy relationships among intestinal bacteria of Aves,
777 Mammalia, and Actinopterygii. Appl Environ Microbiol. 2016;82(5):1496–503.
778 13. Sriswasdi S, Yang CC, Iwasaki W. Generalist species drive microbial dispersion and
779 evolution. Nat Commun [Internet]. 2017;8(1). Available from:
780 http://dx.doi.org/10.1038/s41467-017-01265-1
781 14. Schiro G, Chen Y, Blankinship JC, Barberán A. Ride the dust: linking dust dispersal
782 and spatial distribution of microorganisms across an arid landscape. Environ
783 Microbiol. 2022;24(9):4094–107.
784 15. Parish TR. Surface winds over the Antarctic continent: A review. Rev Geophys
785 [Internet]. 1988;26(1):169. Available from:
786 http://doi.wiley.com/10.1029/RG026i001p00169
787 16. Archer SDJ, Lee KC, Caruso T, Maki T, Lee CK, Cary SC, et al. Airborne microbial
788 transport limitation to isolated Antarctic soil habitats. Nat Microbiol [Internet].
789 2019;4(6):925–32. Available from: http://dx.doi.org/10.1038/s41564-019-0370-4
790 17. Hughes KA, McCartney HA, Lachlan-Cope TA, Pearce DA. A preliminary study of
791 airborne microbial biodiversity over Peninsular Antarctica. Cell Mol Biol (Noisy-le-
792 grand) [Internet]. 2004 Jul;50(5):537–42. Available from:
793 http://www.ncbi.nlm.nih.gov/pubmed/15559970
794 18. Pearce DA, Hughes KA, Lachlan-Cope T, Harangozo SA, Jones AE. Biodiversity of

air-borne microorganisms at Halley station, Antarctica. Extremophiles [Internet]. 2010 Mar 21;14(2):145–59. Available from: http://link.springer.com/10.1007/s00792-009-0293-8

19. Herbold CW, Mcdonald IR, Cary SC. Microbial Ecology of Geothermal Habitats in Antarctica. In: Cowan D, editor. Antarctic Terrestrial Microbiology: Physical and Biological Properties of Antarctic Soils. Berlin, Heidelberg: Springer; 2014. p. 1–328.

20. Mahato NK, Sharma A, Singh Y, Lal R. Comparative metagenomic analyses of a high-altitude Himalayan geothermal spring revealed temperature-constrained habitat-specific microbial community and metabolic dynamics. Arch Microbiol [Internet]. 2019 Apr 25;201(3):377–88. Available from: http://link.springer.com/10.1007/s00203-018-01616-6

21. Weltzer ML, Miller SR. Ecological divergence of a novel group of Chloroflexus strains along a geothermal gradient. Appl Environ Microbiol. 2013;79(4):1353–8.

22. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018;9(1).

23. Power JF, Lowe CL, Carere CR, McDonald IR, Cary SC, Stott MB. Temporal dynamics of geothermal microbial communities in Aotearoa-New Zealand. Front Microbiol [Internet]. 2023 Mar 15;14. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1094311/full

24. Fernandes-Martins MC, Colman DR, Boyd ES. Relationships between fluid mixing, biodiversity, and chemosynthetic primary productivity in Yellowstone hot springs. Environ Microbiol [Internet]. 2023 May 29;25(5):1022–40. Available from: https://sfamjournals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16340

25. Noell SE, Baptista MS, Smith E, McDonald IR, Lee CK, Stott MB, et al. Unique Geothermal Chemistry Shapes Microbial Communities on Mt. Erebus, Antarctica.

Front Microbiol [Internet]. 2022 May 3;13(May). Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2022.836943/full

26. Broady P, Given D, Greenfield L, Thompson K. The biota and environment of fumaroles on Mt Melbourne, northern Victoria Land. Polar Biol. 1987;7(2):97–113.

27. Bargagli R, Broady PA, Walton DWH. Preliminary investigation of the thermal biosystem of Mount Rittmann fumaroles (northern Victoria Land, Antarctica). Antarct Sci. 1996;8(2):121–6.

28. Soo RM, Wood SA, Grzymski JJ, McDonald IR, Cary SC. Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. Environ Microbiol. 2009;11(3):715–28.

29. Herbold CW, Noell SE, Lee CK, Vickers CJ, Stott MB, Eisen JA, et al. Nutritional niches of endemic, aerobic heterotrophs from an isolated Antarctic terrestrial hydrothermal refugium elucidated through metagenomics. Rev. 2024;

30. Skotnicki ML, Selkirk PM, Broady P, Adam KD, Ninham JA. Dispersal of the moss Campylopus pyriformis on geothermal ground near the summits of Mount Erebus and Mount Melbourne, Victoria Land, Antarctica. Antarct Sci [Internet]. 2001 Sep 27;13(3):280–5. Available from: https://www.cambridge.org/core/product/identifier/S0954102001000396/type/journal_article

31. Logan NA, Lebbe L, Hoste B, Goris J, Forsyth G, Heyndrickx M, et al. 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. 2000;50(5):1741–53.

32. Pepi M, Agnorelli C, Bargagli R. Iron demand by thermophilic and mesophilic bacteria isolated from an antarctic geothermal soil. BioMetals. 2005;18(5):529–36.

845 33. Bargagli R, Skotnicki ML, Marri L, Pepi M, Mackenzie A, Agnorelli C. New record of
846 moss and thermophilic bacteria species and physico-chemical properties of geothermal
847 soils on the northwest slope of Mt. Melbourne (Antarctica). Polar Biol.
848 2004;27(7):423–31.
849 34. Park M, Choe Y-H, Kim M. Soil Metagenome Sequences from a Geothermal Site at
850 Mount Melbourne in Antarctica. Microbiol Resour Announc. 2021;10(10):7–8.
851 35. Allan RN, Lebbe L, Heyrman J, De Vos P, Buchanan CJ, Logan NA. Brevibacillus
852 levickii sp. nov. and Aneurinibacillus terranovensis sp. nov., two novel
853 thermoacidophiles isolated from geothermal soils of northern Victoria Land,
854 Antarctica. Int J Syst Evol Microbiol. 2005;55(3):1039–50.
855 36. Poli A, Esposito E, Lama L, Orlando P, Nicolaus G, de Appolonia F, et al.
856 Anoxybacillus amylolyticus sp. nov., a thermophilic amylase producing bacterium
857 isolated from Mount Rittmann (Antarctica). Syst Appl Microbiol. 2006;29(4):300–7.
858 37. Fraser CI, Terauds A, Smellie J, Convey P, Chown SL. Geothermal activity helps life
859 survive glacial cycles. Proc Natl Acad Sci U S A. 2014;111(15):5634–9.
860 38. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3:
861 Fast Estimation of Metagenomic Coverage and Sequence Diversity. Fodor A, editor.
862 mSystems [Internet]. 2018 Jun 26;3(3). Available from:
863 https://journals.asm.org/doi/10.1128/mSystems.00039-18
864 39. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An
865 Expanded Genomic Representation of the Phylum Cyanobacteria. Genome Biol Evol
866 [Internet]. 2014 May;6(5):1031–45. Available from:
867 https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evu073
868 40. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, et al. Genomic
869 and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised

energy source for microbial growth and survival. ISME J [Internet]. 2016 Mar 25;10(3):761–77. Available from: https://www.nature.com/articles/ismej2015153

41. Geyer A. Antarctic volcanism: Active volcanism overview. Geol Soc Mem. 2021;55(1):55–72.

42. Finn DR, Yu J, Ilhan ZE, Fernandes VMC, Penton CR, Krajmalnik-Brown R, et al. MicroNiche: An R package for assessing microbial niche breadth and overlap from amplicon sequencing data. FEMS Microbiol Ecol. 2020;96(8):1–12.

43. von Meijenfeldt FAB, Hogeweg P, Dutilh BE. A social niche breadth score reveals niche range strategies of generalists and specialists. Nat Ecol Evol. 2023;

44. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, et al. Quantifying community assembly processes and identifying features that impose them. ISME J. 2013;7(11):2069–79.

45. DePoy AN, King GM. Distribution and diversity of anaerobic thermophiles and putative anaerobic nickel-dependent carbon monoxide-oxidizing thermophiles in mesothermal soils and sediments. Front Microbiol [Internet]. 2023 Jan 9;13(January):1–14. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1096186/full

46. Zhu J, Ma Y, Huang L, Zhang W. Homogeneous selection is not always important in bacterial community in the eutrophic enclosed bay. Ecol Process [Internet]. 2022 Mar 18;11(1):27. Available from: https://ecologicalprocesses.springeropen.com/articles/10.1186/s13717-022-00373-1

47. Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, et al. Spatial scaling of microbial eukaryote diversity. Nature [Internet]. 2004 Dec;432(7018):747–50. Available from: https://www.nature.com/articles/nature03034

48. Whitaker RJ, Grogan DW, Taylor JW. Geographic Barriers Isolate Endemic

895 Populations of Hyperthermophilic Archaea. Science (80-) [Internet]. 2003 Aug
896 15;301(5635):976–8. Available from:
897 https://www.science.org/doi/10.1126/science.1086909
898 49. Locey KJ, Muscarella ME, Larsen ML, Bray SR, Jones SE, Lennon JT. Dormancy
899 dampens the microbial distance–decay relationship. Philos Trans R Soc B Biol Sci
900 [Internet]. 2020 May 11;375(1798):20190243. Available from:
901 https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0243
902 50. Fang S, An X, Liu H, Cheng Y, Hou N, Feng L, et al. Enzymatic degradation of
903 aliphatic nitriles by Rhodococcus rhodochrous BX2, a versatile nitrile-degrading
904 bacterium. Bioresour Technol [Internet]. 2015 Jun;185:28–34. Available from:
905 https://linkinghub.elsevier.com/retrieve/pii/S0960852415002576
906 51. Houghton KM, Carere CR, Stott MB, McDonald IR. Thermophilic methane oxidation
907 is widespread in Aotearoa-New Zealand geothermal fields. Front Microbiol. 2023
908 Aug;14.
909 52. Op den Camp HJM, Islam T, Stott MB, Harhangi HR, Hynes A, Schouten S, et al.
910 Environmental, genomic and taxonomic perspectives on methanotrophic
911 Verrucomicrobia. Environ Microbiol Rep. 2009 Oct;1(5):293–306.
912 53. Miller SR, Williams C, Strong AL, Carvey D. Ecological Specialization in a Spatially
913 Structured Population of the Thermophilic Cyanobacterium Mastigocladus laminosus.
914 Appl Environ Microbiol [Internet]. 2009 Feb;75(3):729–34. Available from:
915 https://journals.asm.org/doi/10.1128/AEM.01901-08
916 54. Cartwright J, Dzantor EK, Momen B. Soil microbial community profiles and
917 functional diversity in limestone cedar glades. CATENA [Internet]. 2016
918 Dec;147:216–24. Available from:
919 https://linkinghub.elsevier.com/retrieve/pii/S0341816216302697

920 55. Cartwright J, Hui D. Soil respiration patterns and controls in limestone cedar glades.
921 Plant Soil [Internet]. 2015 Apr 7;389(1–2):157–69. Available from:
922 http://link.springer.com/10.1007/s11104-014-2348-6
923 56. Louca S. The rates of global bacterial and archaeal dispersal. ISME J. 2022;16(1):159–
924 67.
925 57. Sharp CE, Brady AL, Sharp GH, Grasby SE, Stott MB, Dunfield PF. Humboldt's spa:
926 Microbial diversity is controlled by temperature in geothermal environments. ISME J.
927 2014;8(6):1166–74.
928 58. Pandit SN, Kolasa J, Cottenie K. Contrasts between habitat generalists and specialists:
929 An empirical extension to the basic metacommunity framework. Ecology.
930 2009;90(8):2253–62.
931 59. Xu Q, Luo G, Guo J, Xiao Y, Zhang F, Guo S, et al. Microbial generalist or specialist:
932 Intraspecific variation and dormancy potential matter. Mol Ecol. 2022;31(1):161–73.
933 60. Wall CA, Koniges GJ, Miller SR. Divergence with gene flow in a population of
934 thermophilic bacteria: a potential role for spatially varying selection. Mol Ecol
935 [Internet]. 2014 Jul;23(14):3371–83. Available from:
936 https://onlinelibrary.wiley.com/doi/10.1111/mec.12812
937 61. He Q, Wang S, Feng K, Michaletz ST, Hou W, Zhang W, et al. High speciation rate of
938 niche specialists in hot springs. ISME J [Internet]. 2023 Jun 7; Available from:
939 https://www.nature.com/articles/s41396-023-01447-4
940 62. Sriaporn C, Campbell KA, Van Kranendonk MJ, Handley KM. Bacterial and archaeal
941 community distributions and cosmopolitanism across physicochemically diverse hot
942 springs. ISME Commun [Internet]. 2023 Aug 18;3(1):80. Available from:
943 https://www.nature.com/articles/s43705-023-00291-z
944 63. Miller SR, Strong AL, Jones KL, Ungerer MC. Bar-coded pyrosequencing reveals

shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. Appl Environ Microbiol. 2009;75(13):4565–72.

64. Büchi L, Vuilleumier S. Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. Am Nat. 2014;183(5):612–24.

65. Wilson DS, Yoshimura J. On the Coexistence of Specialists and Generalists. Am Nat [Internet]. 1994 Oct;144(4):692–707. Available from: https://www.journals.uchicago.edu/doi/10.1086/285702

66. Koffel T, Umemura K, Litchman E, Klausmeier CA. A general framework for species-abundance distributions: Linking traits and dispersal to explain commonness and rarity. Ecol Lett [Internet]. 2022 Nov 14;25(11):2359–71. Available from: https://onlinelibrary.wiley.com/doi/10.1111/ele.14094

67. Marvier M, Kareiva P, Neubert MG. Habitat Destruction, Fragmentation, and Disturbance Promote Invasion by Habitat Generalists in a Multispecies Metapopulation. Risk Anal [Internet]. 2004 Aug 8;24(4):869–78. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.0272-4332.2004.00485.x

68. Odum EP. The Strategy of Ecosystem Development: An understanding of ecological succession provides a basis for resolving man's conflict with nature. Science (80-). 1969;164(3877):262–70.

69. Orrock JL, Watling JI. Local community size mediates ecological drift and competition in metacommunities. Proc R Soc B Biol Sci [Internet]. 2010 Jul 22;277(1691):2185–91. Available from: https://royalsocietypublishing.org/doi/10.1098/rspb.2009.2344

70. Ibarra JT, Martin K. Beyond species richness: an empirical test of top predators as surrogates for functional diversity and endemism. Ecosphere [Internet]. 2015

970 Aug;6(8):art142. Available from: http://doi.wiley.com/10.1890/ES15-00207.1
971 71. Cowling RM, Holmes PM. Endemism and speciation in a lowland flora from the Cape
972 Floristic Region. Biol J Linn Soc [Internet]. 1992 Dec;47(4):367–83. Available from:
973 https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/j.1095-
974 8312.1992.tb00675.x
975 72. Vyverman W, Verleyen E, Wilmotte A, Hodgson DA, Willems A, Peeters K, et al.
976 Evidence for widespread endemism among Antarctic micro-organisms. Polar Sci
977 [Internet]. 2010 Aug;4(2):103–13. Available from:
978 https://linkinghub.elsevier.com/retrieve/pii/S1873965210000149
979 73. Brochier-Armanet C, Gribaldo S, Forterre P. Spotlight on the Thaumarchaeota. ISME
980 J [Internet]. 2012;6(2):227–30. Available from:
981 http://dx.doi.org/10.1038/ismej.2011.145
982 74. Ilanko T, Fischer TP, Kyle P, Curtis A, Lee H, Sano Y. Modification of fumarolic
983 gases by the ice-covered edifice of Erebus volcano, Antarctica. J Volcanol Geotherm
984 Res. 2019;381:119–39.
985 75. Management Plan For Antarctic Specially Protected Area No . 175.
986 2014;13(175):199–225.
987 76. Nieuwenhuize J, Maas YE., Middelburg JJ. Rapid analysis of organic carbon and
988 nitrogen in particulate materials. Mar Chem [Internet]. 1994 Feb;45(3):217–24.
989 Available from: https://linkinghub.elsevier.com/retrieve/pii/0304420394900051
990 77. Caporaso JG, Ackermann G, Apprill A, Bauer M, Berg-Lyons D, Betley J, et al. Earth
991 microbiome project: EMP 16S Illumina Amplicon Protocol. ProtocolsIo [Internet].
992 2018;1–7. Available from: https://dx.doi.org/10.17504/protocols.io.nuudeww
993 78. Ul-Hasan S, Bowers RM, Figueroa-Montiel A, Licea-Navarro AF, Beman JM, Woyke
994 T, et al. Community ecology across bacteria, archaea and microbial eukaryotes in the

995 sediment and seawater of coastal Puerto Nuevo, Baja California. Brusetti L, editor.
996 PLoS One [Internet]. 2019 Feb 14;14(2):e0212355. Available from:
997 https://dx.plos.org/10.1371/journal.pone.0212355
998 79. Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, et al. Microbial 16S
999 rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM)
1000 Platform. J Microbiol Methods [Internet]. 2012;91(1):80–8. Available from:
1001 http://dx.doi.org/10.1016/j.mimet.2012.07.008
1002 80. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2:
1003 High-resolution sample inference from Illumina amplicon data. Nat Methods.
1004 2016;13(7):581–3.
1005 81. Wright ES. Using DECIPHER v2.0 to analyze big biological sequence data in R. R J.
1006 2016;8(1):352–9.
1007 82. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics [Internet]. 2011 Feb
1008 15;27(4):592–3. Available from: https://academic.oup.com/bioinformatics/article-
1009 lookup/doi/10.1093/bioinformatics/btq706
1010 83. Andrews S. FastQC [Internet]. Available from:
1011 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
1012 84. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for
1013 multiple tools and samples in a single report. Bioinformatics [Internet]. 2016 Oct
1014 1;32(19):3047–8. Available from:
1015 https://academic.oup.com/bioinformatics/article/32/19/3047/2196507
1016 85. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
1017 sequence data. Bioinformatics [Internet]. 2014 Aug 1;30(15):2114–20. Available from:
1018 https://academic.oup.com/bioinformatics/article/30/15/2114/2390096
1019 86. Zhang Z, Zhang L, Zhao Z, Wang H, Ju F. Deduplication Improves Cost-Efficiency

1020 and Yields of De novo Assembly and Binning of Shot-Gun
1021 Metagenomes in Microbiome Research. bioRxiv [Internet]. 2022 Jan
1022 1;2022.10.12.512008. Available from:
1023 http://biorxiv.org/content/early/2022/10/17/2022.10.12.512008.abstract
1024 87. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
1025 Bioinformatics [Internet]. 2018 Sep 1;34(17):i884–90. Available from:
1026 https://academic.oup.com/bioinformatics/article/34/17/i884/5093234
1027 88. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets.
1028 Bioinformatics [Internet]. 2011 Mar 15;27(6):863–4. Available from:
1029 https://academic.oup.com/bioinformatics/article/27/6/863/236283
1030 89. Rodriguez-R LM, Konstantinidis KT. Nonpareil: a redundancy-based approach to
1031 assess the level of coverage in metagenomic datasets. Bioinformatics [Internet]. 2014
1032 Mar 1;30(5):629–35. Available from:
1033 https://academic.oup.com/bioinformatics/article/30/5/629/246142
1034 90. Uritskiy G V., DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-
1035 resolved metagenomic data analysis. Microbiome [Internet]. 2018 Dec 15;6(1):158.
1036 Available from:
1037 https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0541-1
1038 91. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. Gene and translation initiation site
1039 prediction in metagenomic sequences. Bioinformatics [Internet]. 2012 Sep
1040 1;28(17):2223–30. Available from: https://academic.oup.com/bioinformatics/article-
1041 lookup/doi/10.1093/bioinformatics/bts429
1042 92. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
1043 Prokaryotic gene recognition and translation initiation site identification. BMC
1044 Bioinformatics. 2010;11.

1045 93. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive
1046 binning algorithm for robust and efficient genome reconstruction from metagenome
1047 assemblies. PeerJ [Internet]. 2019 Jul 26;7:e7359. Available from:
1048 https://peerj.com/articles/7359
1049 94. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to
1050 recover genomes from multiple metagenomic datasets. Bioinformatics [Internet]. 2016
1051 Feb 15;32(4):605–7. Available from:
1052 https://academic.oup.com/bioinformatics/article/32/4/605/1744462
1053 95. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
1054 metagenomic contigs by coverage and composition. Nat Methods [Internet]. 2014 Nov
1055 14;11(11):1144–6. Available from: http://www.nature.com/articles/nmeth.3103
1056 96. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from
1057 uncultivated genomes of the global human gut microbiome. Nature [Internet]. 2019
1058 Apr 13;568(7753):505–10. Available from: http://www.nature.com/articles/s41586-
1059 019-1058-x
1060 97. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing
1061 the quality of microbial genomes recovered from isolates, single cells, and
1062 metagenomes. Genome Res [Internet]. 2015 Jul;25(7):1043–55. Available from:
1063 http://genome.cshlp.org/lookup/doi/10.1101/gr.186072.114
1064 98. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly
1065 classification with the genome taxonomy database. Borgwardt K, editor.
1066 Bioinformatics [Internet]. 2022 Nov 30;38(23):5315–6. Available from:
1067 https://academic.oup.com/bioinformatics/article/38/23/5315/6758240
1068 99. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics [Internet].
1069 2014 Jul 15;30(14):2068–9. Available from:

1070 https://academic.oup.com/bioinformatics/article/30/14/2068/2390517
1071 100. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from
1072 cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci
1073 U S A. 2021;118(12):1–10.
1074 101. Li G, Rabe KS, Nielsen J, Engqvist MKM. Machine Learning Applied to Predicting
1075 Microorganism Growth Temperatures and Enzyme Catalytic Optima. ACS Synth Biol.
1076 2019;8(6):1411–20.
1077 102. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. J Open
1078 Source Softw [Internet]. 2016 Sep 14;1(5):27. Available from:
1079 http://joss.theoj.org/papers/10.21105/joss.00027
1080 103. Koslicki D, White S, Ma C, Novikov A. YACHT: an ANI-based statistical test to
1081 detect microbial presence/absence in a metagenomic sample. bioRxiv Prepr Serv Biol
1082 [Internet]. 2023 Apr 20; Available from:
1083 http://www.ncbi.nlm.nih.gov/pubmed/37131762
1084 104. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
1085 Community Ecology Package [Internet]. 2020. Available from: https://cran.r-
1086 project.org/package=vegan
1087 105. Gao C. ggVennDiagram: A “ggplot2” Implement of Venn Diagram [Internet]. 2021.
1088 Available from: https://cran.r-project.org/package=ggVennDiagram
1089 106. Kolde R. pheatmap: Pretty Heatmaps [Internet]. 2019. Available from: https://cran.r-
1090 project.org/package=pheatmap
1091 107. Liu C, Cui Y, Li X, Yao M. microeco : an R package for data mining in microbial
1092 community ecology. FEMS Microbiol Ecol [Internet]. 2021 Jan 26;97(2). Available
1093 from: https://academic.oup.com/femsec/article/doi/10.1093/femsec/fiaa255/6041020
1094 108. Lahti L, Shetty S. Tools for microbiome analysis in R [Internet]. Bioconductor; 2017.

Available from: http://microbiome.github.com/microbiome

109. Greening C, Geier R, Wang C, Woods LC, Morales SE, McDonald MJ, et al. Diverse hydrogen production and consumption pathways influence methane production in ruminants. ISME J [Internet]. 2019 Oct 26;13(10):2617–32. Available from: https://www.nature.com/articles/s41396-019-0464-2

110. Liao H, Liu C, Ai C, Gao T, Yang Q-E, Yu Z, et al. Mesophilic and thermophilic viruses are associated with nutrient cycling during hyperthermophilic composting. ISME J [Internet]. 2023 Jun 8;17(6):916–30. Available from: https://www.nature.com/articles/s41396-023-01404-1

111. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ [Internet]. 2014 Jan 9;2:e243. Available from: https://peerj.com/articles/243

112. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Ponty Y, editor. Bioinformatics [Internet]. 2019 Oct 15;35(20):4162–4. Available from: https://academic.oup.com/bioinformatics/article/35/20/4162/5378708

113. Hernandez DJ, Kiesewetter KN, Almeida BK, Revillini D, Afkhami ME. Multidimensional specialization and generalization are pervasive in soil prokaryotes. Nat Ecol Evol [Internet]. 2023 Aug 7; Available from: https://doi.org/10.1101/2023.03.17.533171

114. Hijmans RJ. geosphere: Spherical Trigonometry [Internet]. 2022. Available from: https://cran.r-project.org/package=geosphere

115. Chao A, Chazdon RL, Colwell RK, Shen T-J. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol Lett [Internet]. 2004 Dec 17;8(2):148–59. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1461-0248.2004.00707.x

116. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome [Internet]. 2022;10(1):1–22. Available from: https://doi.org/10.1186/s40168-021-01213-8

117. R Core Team. R: A language and environment for statistical computing [Internet]. 2020. Available from: https://www.r-project.org/

118. Wickham H. Ggplot2: Elegant Graphics For Data Analysis [Internet]. New York: Springer-Verlag; 2016. Available from: https://ggplot2.tidyverse.org

119. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw [Internet]. 2019 Nov 21;4(43):1686. Available from: https://joss.theoj.org/papers/10.21105/joss.01686

120. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots [Internet]. 2023. Available from: https://cran.r-project.org/package=ggpubr

1134 **Figure Legends**

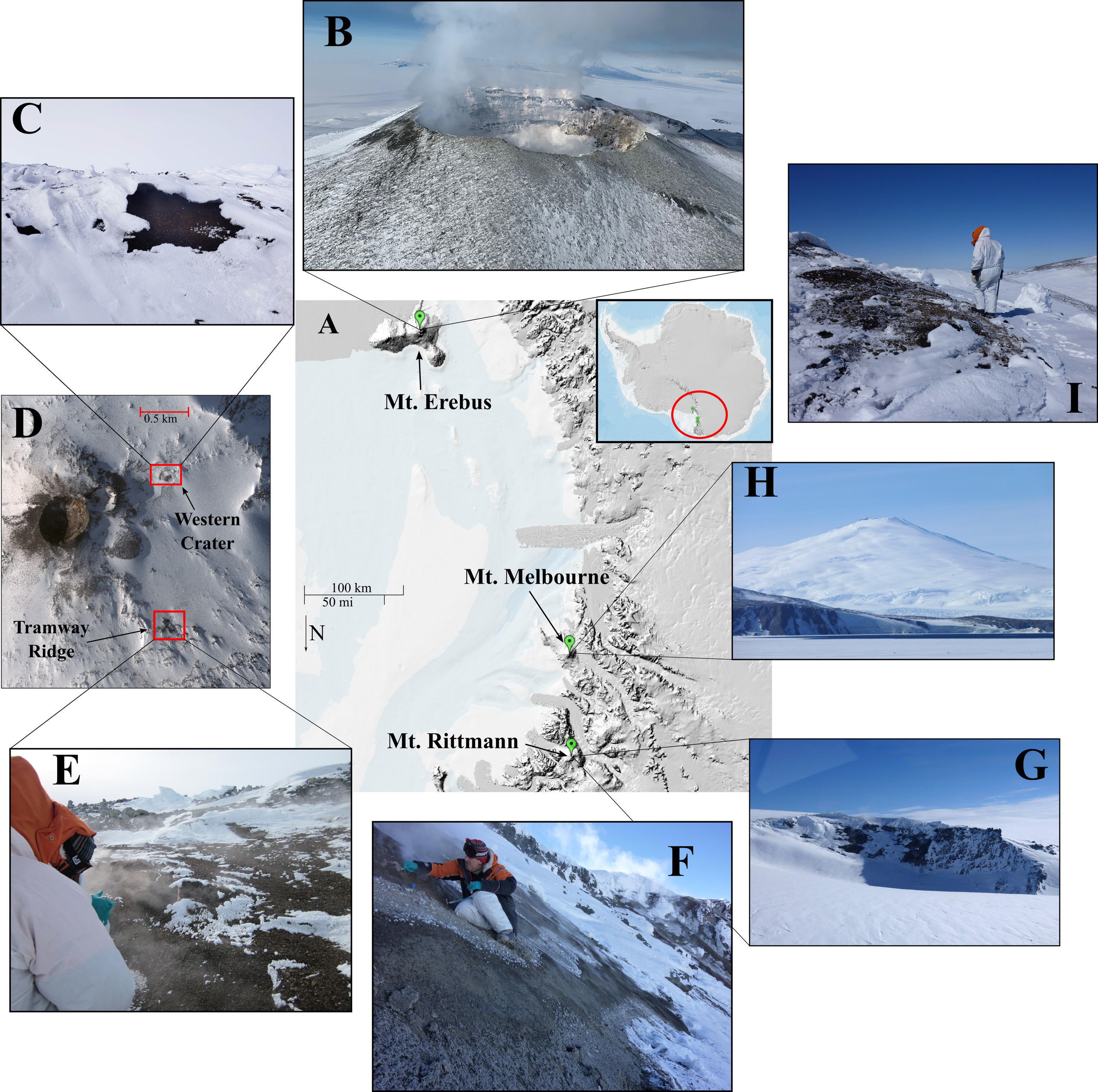
1135 Figure 1 Map and images of sampling sites. (A) Satellite image of Antarctica (top right inset,
1136 red circle indicates Victoria Land); main image shows Victoria Land and Ross Island. (B)
1137 Aerial picture of Mt. Erebus main crater. (C) Picture of the hot soil site sampled at Western
1138 Crater on Mt. Erebus. (D) Satellite image showing the summit of Mt. Erebus and the two
1139 sampling sites in this study, Tramway Ridge (-77.519444, 167.116389 DD) and Western
1140 Crater (-77.519167, 167.187778 DD). (E) Image of sampling site on Tramway Ridge. (F)
1141 Image of sampling on the side of Mt. Rittmann (-73.448445, 165.497444 DD). (G) Image of
1142 Mt. Rittmann. (H) Image of Mt. Melbourne. (I) Image of sampling at a hot soil site on Mt.
1143 Melbourne (-74.350016, 164.700139 DD). Satellite image in (A) is from the Antarctic Digital
1144 Database Map Viewer https://www.add.scar.org/, Open Source, using hillshade and
1145 bathymetry layers. Satellite image in (D) was purchased from DigitalGlobe Incorporated,
1146 Longmont CO USA (2019). All other images were taken by Ian McDonald and S. Craig
1147 Cary.

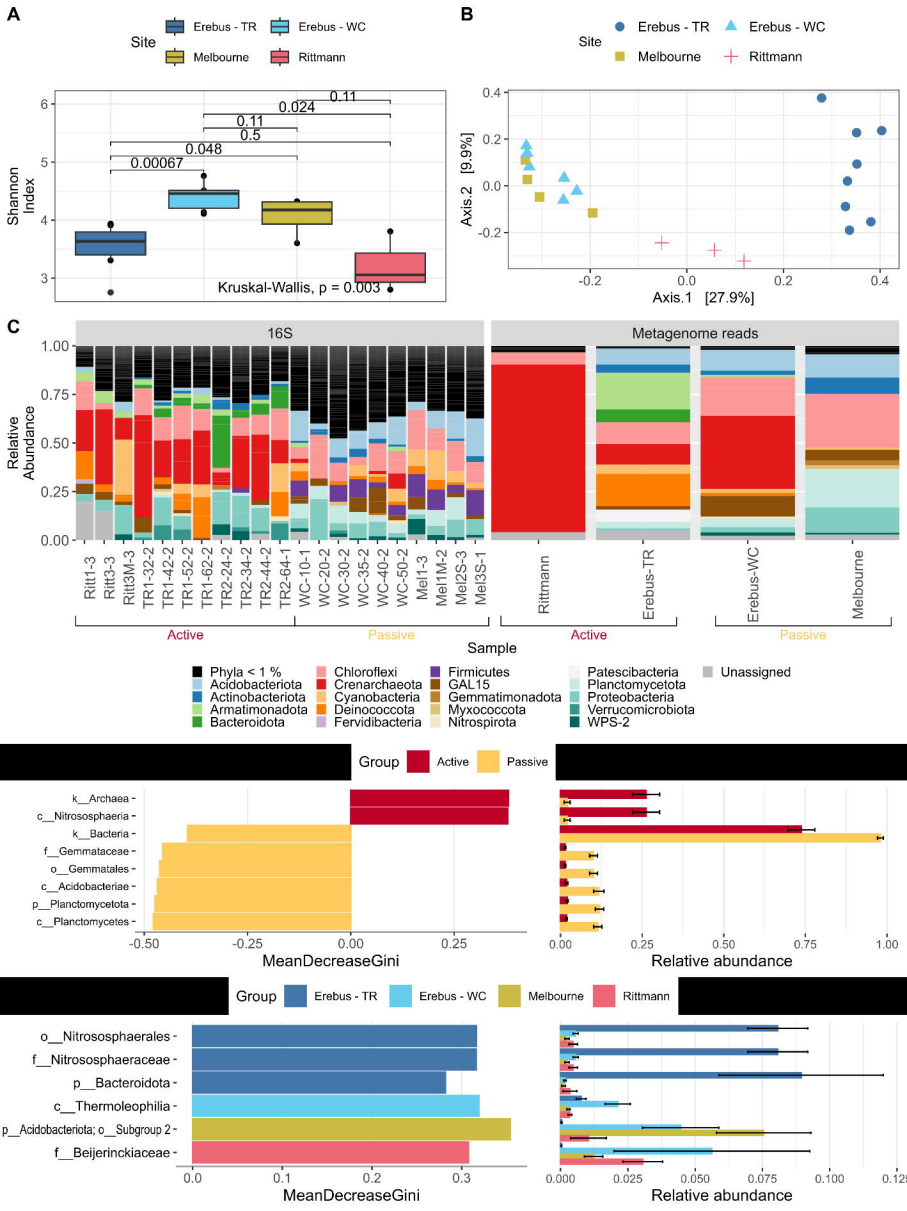
1148 Figure 2 Diversity and composition of microbial communities across the four sites. (A) Alpha
1149 diversity of microbial communities based on 16S rRNA gene sequences across the four sites
1150 (Tramway Ridge (TR), Western Crater (WC), Melbourne, or Rittmann, n = 8, 6, 4, or 3
1151 respectively) using the Shannon index of diversity. Paired significance test used an
1152 undirected Wilcox test. (B) Principal coordinates analysis (PCoA) of the microbial
1153 communities based on Unifrac similarities between 16S rRNA gene sequences. The
1154 percentages in brackets for each axis indicate the percentage of the variation in community
1155 composition explained along that axis. (C) Microbial community composition at the phylum
1156 level for both 16S rRNA gene sequencing samples (16S) and metagenome samples at the
1157 read level (Metagenome reads); singleM was used for determining taxonomy for the
1158 metagenome reads. For the 16S rRNA gene sequence data, the sequencing run used for

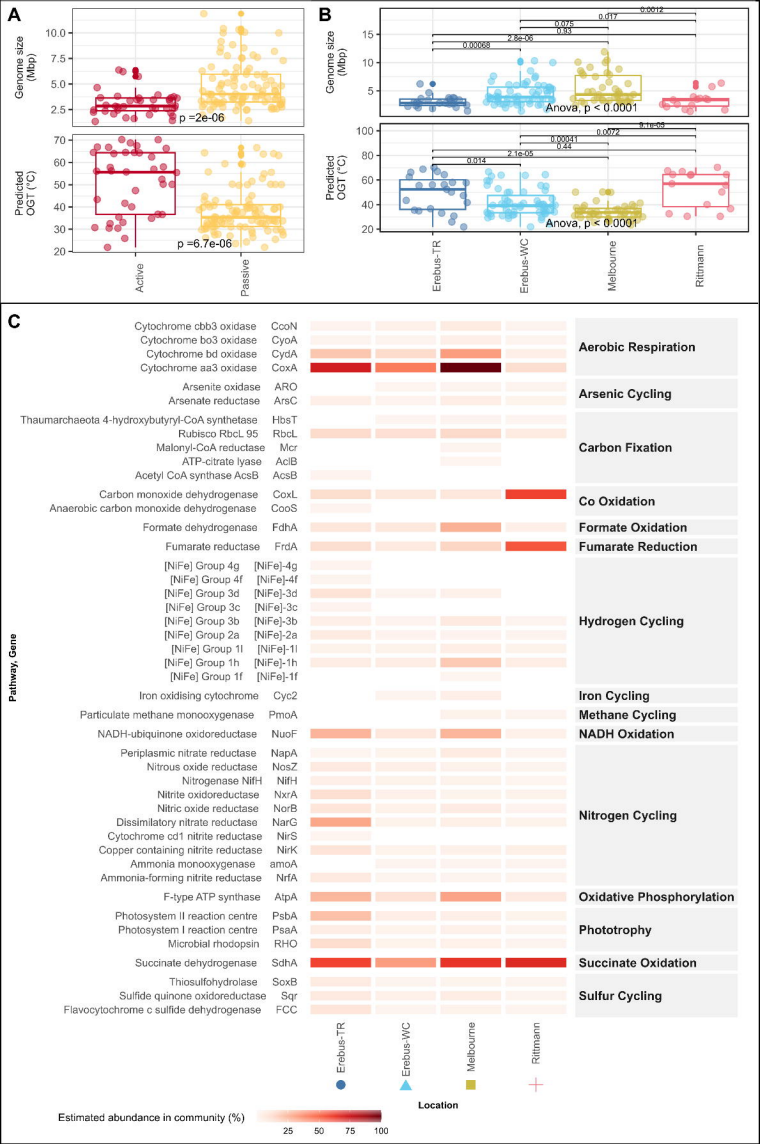
1159 Tramway Ridge (TR) and Western Crater (WC) samples are denoted by the last digit, either 1
1160 for the first (data from (25)) or second sequencing run. For this plot, phylum names for
1161 metagenome reads were changed to match SILVA 16S rRNA gene taxonomy to allow for
1162 ease of comparison between the two methods (*Thermoproteota* to *Crenarchaeota*,
1163 *Chloroflexota* to *Chloroflexi*, CSP1-3 to GAL15, *Dormibacterota* to *Chloroflexi* (class AD3),
1164 *Eremiobacterota* to WPS-2). (D, E) Results of a differential abundance random forest
1165 analysis to identify taxa at any level that are most important for distinguishing between (D)
1166 active and passive sites, or (E) all four sites in the 16S rRNA gene sequence data. The plots
1167 on the left give the mean decrease in Gini value, which is a measure of the importance of that
1168 taxa to distinguishing the sites. The plot on the right shows the relative abundance of those
1169 taxa at the respective sites. A threshold of either 0.2 or 0.1 relative abundance was used for
1170 plotting ASVs for the active vs. passive comparison or all four sites comparison, respectively.
1171 Figure 3 Genomic and functional characteristics of MAGs and metagenomes from each site.
1172 (A, B) Comparison of genome characteristics for MAGs recovered from different site
1173 categories (A) or sites (B). (A) n = 32 or 79 for active or passive, respectively (*p*-values: two-
1174 sided Wilcox test). (B) n = 19, 37, 42, or 13 for Tramway Ridge (TR), Western Crater (WC),
1175 Melbourne, or Rittmann, respectively. p-values given for the pairwise comparisons are for
1176 two-sided Wilcox tests. (C) Analysis of the functional potential of the microbial communities
1177 at the four site samples with metagenomes, based on the abundance of marker genes for
1178 different metabolic pathways. Abundance within community was calculated using RPKM
1179 values, which were normalized by the abundance of single copy marker genes within each
1180 metagenome. TR: Tramway Ridge. WC: Western Crater.
1181 Figure 4 Analysis of Antarctic geothermal *Cyanobacteria* diversity and factors shaping it.
1182 Abundance heatmaps of *Cyanobacteria* communities using (A) 16S rRNA gene sequence and
1183 (B) metagenome assembled genome (MAG) data. For (A), relative abundances of

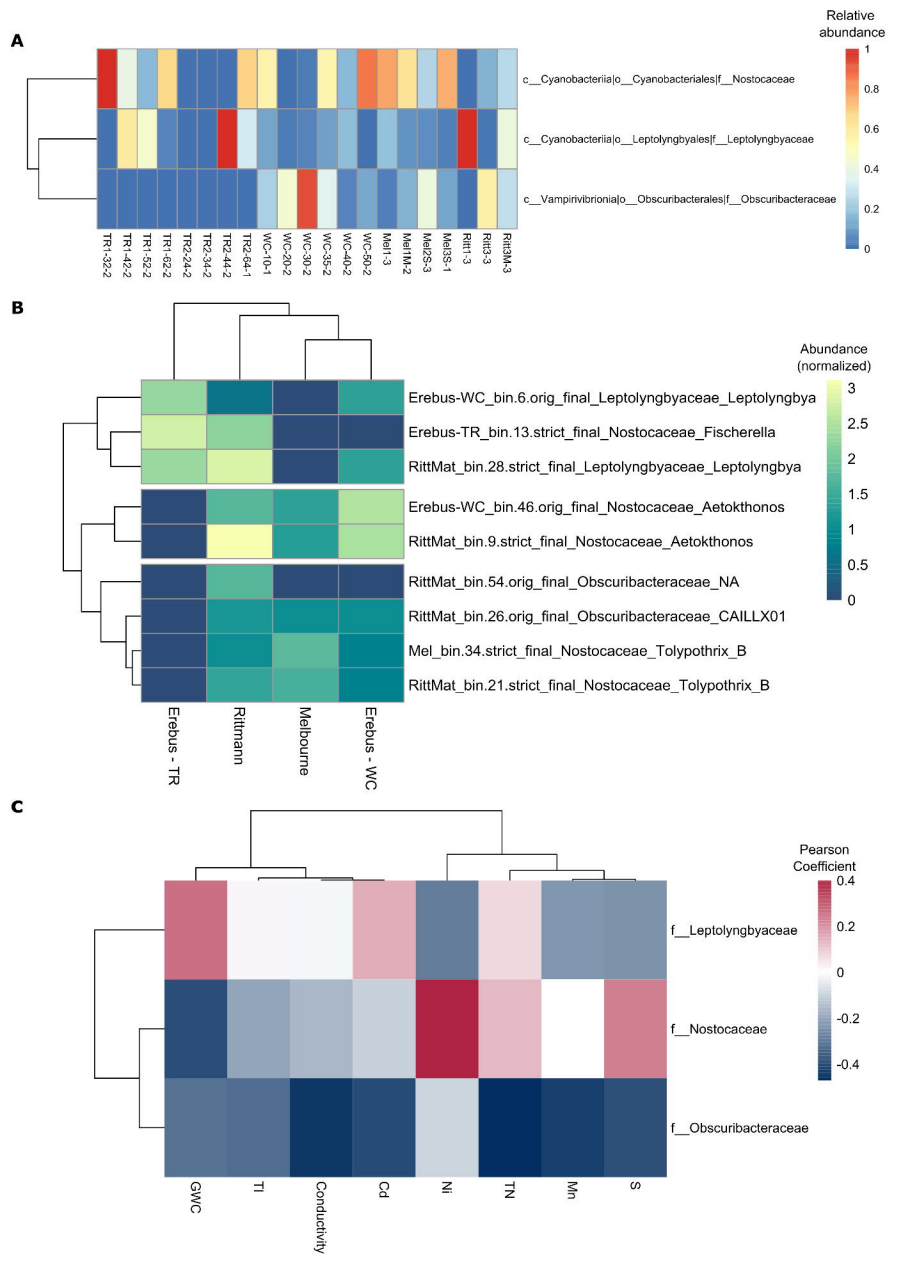
1184 *Cyanobacteria* families are presented as relative to the abundance of other Cyanobacteria.
1185 For (B), MAG abundances were marked as zero if the MAG was not present at the site. Then,
1186 the abundances were normalized: $\log_{10}(\text{abundance} + 0.001) + 4$. (C) Correlation heatmap,
1187 using the Pearson coefficient, of the cyanobacterial families to the relevant environmental
1188 variables (selected for having Mantel test $p < 0.05$). GWC: gravitational water content; TN:
1189 total nitrogen.
1190 Figure 5 Ecological predictions for microorganisms at Antarctic geothermal sites. (A-C)
1191 Frequency plots identifying the number of either specialists or generalists for (A) habitat
1192 (using Levin's niche breadth, B_n), (B) temperature (using Hulbert's niche breadth, B_n), or (C)
1193 pH (using Hulbert's niche breadth, B_n) in the 16S rRNA gene amplicon sequence data.
1194 Arrows point to generalists where difficult to visualize. (D) Contribution of different
1195 ecological factors to shaping the community assembly of microorganisms in the 16S rRNA
1196 gene amplicon data. (E, F, G) Distance-decay plots for the 16S rRNA gene amplicon data
1197 with a linear regression best fit model for (E) all samples, (F) only samples from Active sites,
1198 or (G) only samples from Passive sites.
1199 Figure 6 (A-B) Exploring evidence for genomic markers of metabolic generalism in
1200 metagenome assembled genomes (MAGs) from the four sites. For (A), n = 35, 52, 27, or 2
1201 for the categories 1 site, 2 sites, 3 sites, or 4 sites, respectively. p-values for ANOVA tests
1202 conducted in (A) are, from top left clockwise order: 0.08, 0.07, 0.04, and 0.02. p-value shown
1203 on plot is the result of an undirect Wilcox test, Benjamini-Hochburg corrected. No
1204 differences for pfam domains remained significant after BH correction. For (B) and (C), n =
1205 11, 2, or 7 for Tramway Ridge (TR) only, Rittmann only, or TR/Rittmann only MAGs,
1206 respectively; n = 5, 3, 8, 15, or 2 for TR/Western Crater (WC) only, WC/Rittmann only,
1207 TR/WC/Rittmann only, WC/Melbourne/Rittmann only, or MAGs at all four sites,
1208 respectively; n = 10, 11, or 36 for WC only, Melbourne only, or WC/Melbourne only MAGs,

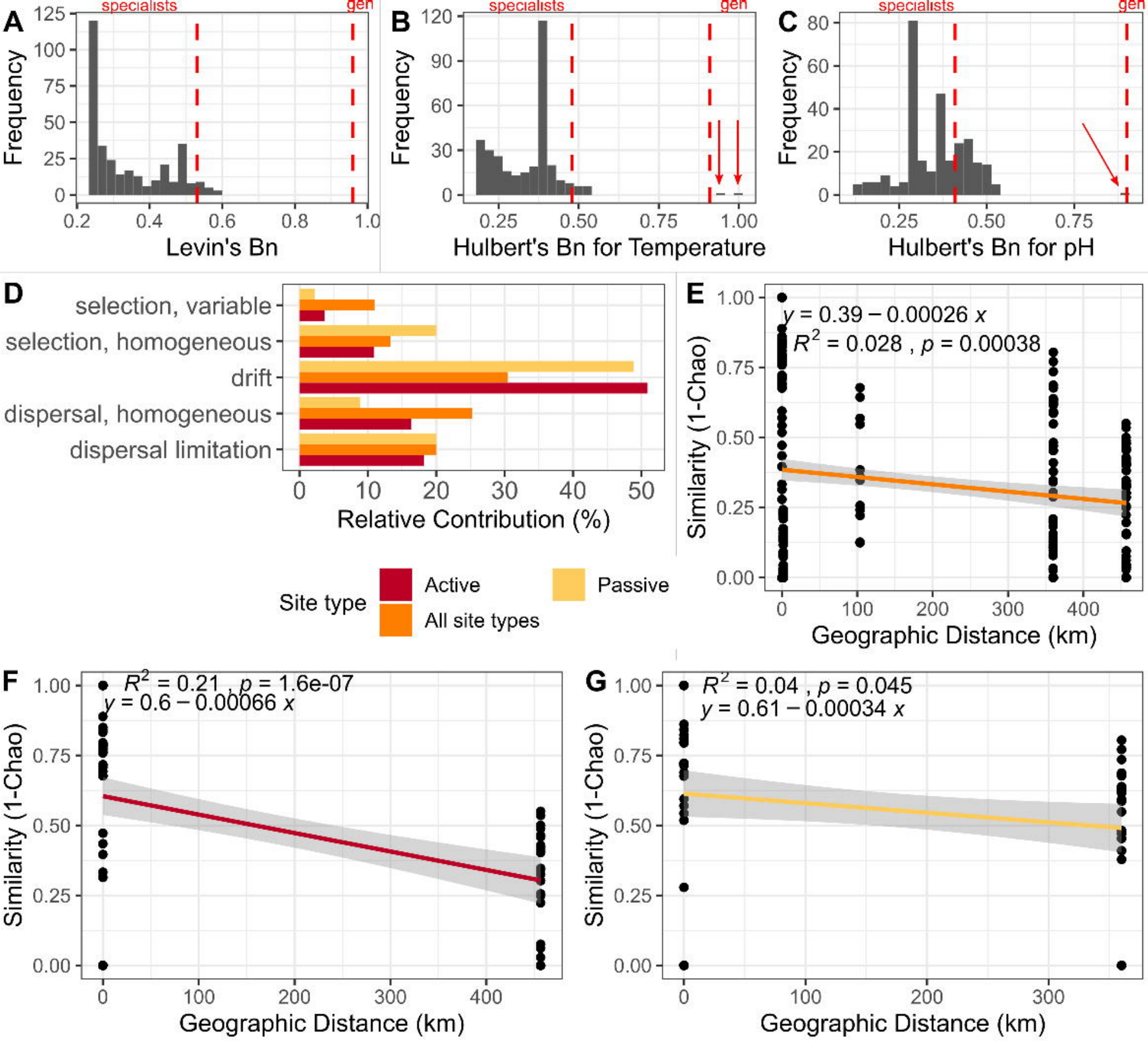
respectively. For (B), * represents a statistically significant ($p < 0.05$) result of an undirected Wilcox test after Benjamini-Hochburg correction. See Methods for how Metabolic Pathways/Categories were marked as present or absent in MAGs. (C) Heatmap showing functional predictions for MAGs based on 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.

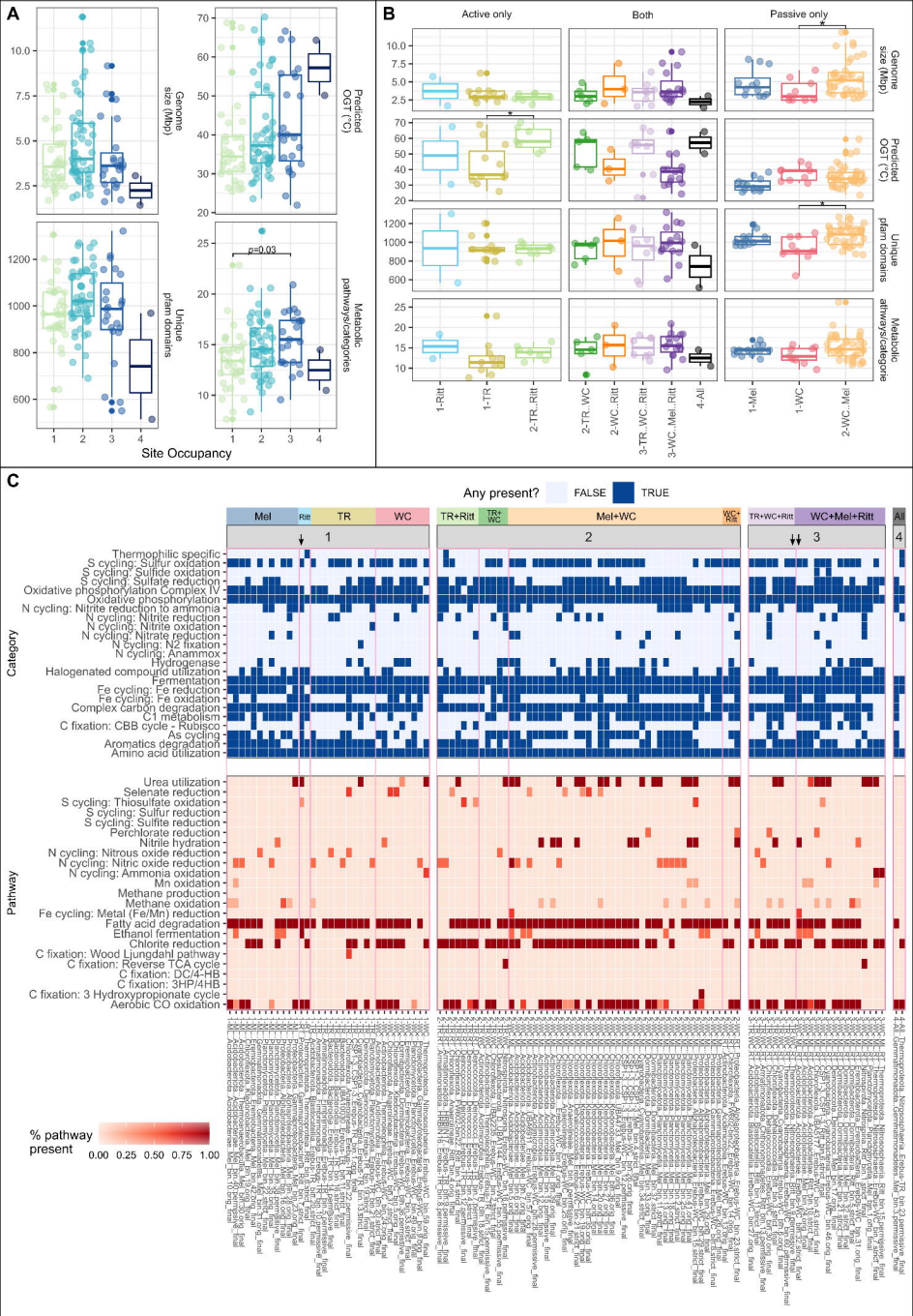

A
B
C
D
E
F
G
H
I
Mt. Erebus
Mt. Melbourne
Mt. Rittmann
100 km
50 mi
N
0.5 km
Western
Crater
Tramway
Ridge


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 (Mbp)
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
0.44
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 4f [NiFe]-4f
[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
Cytochrome cd1 nitrite reductase NirS
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