

Depth-specific distribution of bacterial MAGs in permafrost active layer in Ny Ålesund, Svalbard (79°N)

Katie Sipes^{a*#}, Joy Buongiorno^a, Andrew D. Steen^{a,b}, Andrey A. Abramov^c, Chukwufumnanya Abuah^a, Samantha L. Peters^d, Richard J. Gianonne^d, Robert L. Hettich^d, Julia Boike^{e,f}, Sarahi L. Garcia^{g,h}, Tatiana A. Vishnivetskaya^a, Karen G. Lloyd^a

^a=Department of Microbiology, University of Tennessee, Knoxville, United States

^bDepartment of Earth and Planetary Sciences, University of Tennessee, Knoxville, United States

^cSoil Cryology Laboratory, Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino, Russia

^dBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

^eAlfred Wegener Institute for Polar and Marine Research, Potsdam, Germany

^fDepartment of Geography, Humboldt University, Berlin, Germany

^gDepartment of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

^h Institute for Chemistry and Biology of the Marine Environment (ICBM), School of Mathematics and Science, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

*Current affiliation: Department of Environmental Sciences, Aarhus University, 4000 Roskilde, Denmark

Corresponding author: ksipes@envs.au.dk

Running title: Depth distribution of MAGs in Ny Ålesund, Svalbard

Abstract:

Arctic soil microbial communities may shift due to increasing temperatures and water availability from climate change. We examined temperature and volumetric liquid water content (VWC) in the upper 80 cm of permafrost-affected soil over 2 years (2018-2019) at the Bayelva monitoring station, Ny Ålesund, Svalbard. We found that VWC increases with depth, whereas *in situ* temperature is more stable vertically, ranging from -5°C to 5°C seasonally. Metagenome-assembled genomes (MAGs) of Bacteria and Archaea were obtained at 2-4 cm vertical resolution collected while frozen in April 2018 and at 10 cm vertical resolution collected while thawed in September 2019. The most abundant MAGs affiliated with *Acidobacteriota*, *Actinomycetota*, and *Chloroflexota*. *Actinomycetota* and *Chloroflexota* gradually increase with depth, while *Acidobacteriota* classes *Thermoanaerobaculia* Gp7-AA8, *Blastocatellia* UBA7656, and *Vicinamibacteria* *Vicinamibacterales* are only found above 6 cm, below 6 cm, and below 20 cm, respectively. All MAGs have diverse carbon-degrading genes, and *Actinomycetota* and *Chloroflexota* have autotrophic genes. Genes encoding β -glucosidase, N-acetyl- β -D-glucosaminidase, and xylosidase increase with depth, indicating a greater potential for organic matter degradation with higher VWC. *Acidobacteriota* dominate the top 6 cm with their classes segregating by depth, whereas *Actinomycetota* and *Chloroflexota* dominate below ~ 6 cm and their classes are not segregated by depth. This suggests that *Acidobacteriota* classes adapt to lower VWC at the surface, while *Actinomycetota* and *Chloroflexota* persist better below 6 cm with higher VWC. This indicates that VWC may be as important as temperature in microbial climate change responses in Arctic mineral soils. Here we describe MAG-based Seqcode type species in the *Acidobacteriota*, *Onstottus arcticum* and *Onstottus frigus* in the novel order *Onstottales* and novel family *Onstottaceae*, and *Gilichinskyi gelida* in the novel order *Gilichinskyiales* and novel family *Gilichinskyiaceae*. In the *Actinobacteriota*, we describe *Mayfieldus profundus* in the novel family *Mayfieldaceae*.

Importance:

The high Arctic is warming rapidly. Understanding the fate of organic matter is crucial for the permafrost-thawing based greenhouse effect. Soil microbes convert organic matter to carbon dioxide (CO₂) or methane (CH₄). We examined these microbes in Svalbard's permafrost-affected soils with high depth resolution and considered liquid water availability. Our results show organic matter-degrading enzymes in deeper soils, suggesting they can break down complex carbon compounds. These microbial populations respond quickly to seasonal changes in soil temperature and water content, indicating that future fluctuations driven by Arctic warming could significantly impact microbial community shifts.

Introduction

The high Arctic archipelago, Svalbard, experiences amplified effects of climate change because it receives warm water from the Atlantic Ocean and warm air from greenhouse gas amplification at the pole [18]. Temperatures on Svalbard are often close to the freezing point, especially in the vicinity of the coasts in the southern archipelago, hence soil has been observed to be warming since observations began in 1998 [5,10,48]. Air temperature warming on Svalbard has been even stronger than the Arctic average, with warming rates of about 1.5°C per decade from 1991-2020 [20,41] linked to strong sea-ice decline in that region. However, to fully assess the thermal status of the soil, volumetric liquid water content (VWC) measurements are required in

addition to temperatures. Geological models investigating the impact of soil destabilization on the greenhouse gas budget commonly incorporate the microbial degradation of soil organic matter [1,23,62]. Permafrost temperatures on Svalbard are highest in the vicinity of the coasts and often are close to the freezing point [25], so most of the energy from temperature changes goes into phase changes of water [5]. Therefore, microbial communities will also experience climate change as having access to liquid water more frequently, over longer time intervals, and in greater volumes. VWC is also related to osmolarity in the frozen soils, since water excludes solutes when it freezes, leaving behind higher osmolarity liquid water, which can have an effect on the microbial community [52]. Given that seasonal freezing and thawing are transmitted vertically, microbial reactions to the increased access to liquid water are likely to be depth-dependent over the scale of centimeters. These soils contain diverse microbial populations capable of degrading a wide range of carbon substrates, shown through enzyme activity measurements, metagenomics, and microbial isolates [56,58]. In this study, we investigated the changes in microbial communities responsible for soil organic matter degradation in fine-scale intervals of Svalbard active layer soils using metagenomes, temperature, and VWC.

Previous bacterial taxonomy studies in Arctic permafrost-affected soils, including those from Svalbard, have focused on the upper surface (1-5 cm [21]), and deeper horizons of the active layer (e.g., <20 cm [52,69], 30 cm [32,71], 40 cm [38], and 60 cm [19,71]). All these studies have found microbial communities dominated by *Actinomycetota*, *Acidobacteriota* and *Chloroflexota* which contain key carbon degrading genes [66], establishing them as important players in the microbial response to global climate change in mineral Arctic soils. However, little is known about how microbial communities vary over the small spatial scales that are relevant to these vertically layered and horizontally heterogeneous soils. Since physicochemical characteristics change with depth within the active layer [5], different microbial communities may be adapted to unique conditions at different depth layers. To fill this knowledge gap, the microbial community and its metabolic properties need to be investigated at a fine depth scale [13]. Understanding these depth-dependent variations in microbial degradation of soil organic carbon in the active layer is important for determining the response of these communities to climate change [77]. The redistribution of thawed carbon-rich soil into a cold unfrozen layer near the permafrost could either slow down [4] or speed up [65] microbial activity and organic matter degradation.

We conducted a study in the upper layer of permafrost affected soils at five sites in the Bayelva site, near Ny Ålesund, Svalbard, located on the Leirhaugen hill, 2 km north of the rapidly retreating tongue of Austre Brøggerbreen glacier (Figure 1). The Leirhaugen hill is covered with permafrost patterned ground and sparse vegetation. The annual physical characteristics of soil and snow at the Bayelva site have been continuously monitored since 1998 [5]. To improve understanding of the microbial communities from this poorly characterized environment, frozen winter cores collected in April 2018 were subsampled at high stratigraphic resolution (1 cm to 2 cm depth intervals). To test the hypothesis that the microbial community changes according to depth variations throughout the year, thawed summer pit samples were obtained in early September 2019 and the same measurements were made.

Studying microorganisms within permafrost-affected soils is difficult for many reasons; first, low microbial biomass results in negligible DNA concentrations, which is troublesome for getting objective sequencing data; second, constrained cultivation approaches lead to poor species diversity; and third, rapidly changing environmental characteristics make it difficult to make broadly representative measurements of microbial activity [22]. Amplicon studies of 16S rRNA genes are heavily biased toward known microorganisms [29] because the relative abundance of

microorganisms is often skewed due to primer bias [61]. Therefore, we used metagenomic approaches to analyze microbial DNA [67]. The bulk genomic DNA and the metagenomic-assembled genomes (MAGs) were analyzed to assess the relative abundance of different taxonomic groups and their genetic material for metabolic potential. MAGs are a reliable method to assign functions at lower taxonomic groups across environmental microbial communities, however, they are often difficult to assemble and bin from soils samples due to high microbial diversity [32]. In our approach, we 1) binned MAGs separately in 56 similar soil metagenomic libraries, increasing the chances of binning microorganisms endemic to each horizon, 2) recruited metagenomic reads from each of the 56 soil samples to the full set of MAGs that were produced from all the samples so that a MAG's distribution across samples could be determined even if it was not directly binned in all soil samples, and 3) determined taxonomic distributions across the total metagenomic assemblies to determine if MAG diversity is representative of the microbial community.

Results

Soil physical processes and properties

The quality-controlled dataset collected in the late winter 2018 and summer 2019 from the Bayelva high Arctic permafrost long-term observation site revealed that the onset of freezing, defined as temperatures dropping below zero, started from the top down and was delayed toward the deeper layers (Figure 2 and Table 1). The water in the upper layer soils froze first, before the layers below experienced a decrease in temperature. Consequently, the lower layers of soil reached subzero temperatures only after the upper layers froze permanently for the winter season. Due to the proximity of temperatures of the permafrost affected soils to the freezing point of water (within the soil freezing characteristic curve [18]), the vertical temperature transmission is buffered by phase changes between ice and liquid water.

The duration of the zero-curtain period (the time when the soils is below 0°C and remain there until spring) varied between 18 to 35 days among the soil layers, indicating that differences in soil texture, properties and volumetric liquid water content (VWC) influenced the soil's ability to freeze uniformly. During the winter when the frozen cores were drilled, the VWC in the frozen soil varied between layers, ranging from ~4% to 10% with similar mean temperatures between -8 to -9°C. This variation could be attributed to different soil properties, such as the amount of organic matter and mineral content.

Description of metagenome-assembled genomes

A total of 56 samples collected across five sites in the Bayelva plain in the Laugerhain glacier moraine were sequenced on Illumina NovaSeq, totaling 1.5 TB of metagenomic reads (Table S1) and 169 MAGs of medium to high quality, based on the MIMAGs specification, with $\geq 50\%$ completeness and $\leq 10\%$ contamination (Figure S1, Table S2). Of the 56 metagenomes assembled, 52 yielded MAGs; exceptions were BPF2 (10-12 cm, 18-20 cm), BPF4 (1-2 cm), and BPF5 (43-45 cm), likely due to their smaller assembly sizes. Individual MAGs' assembly size, number of MAGs, or MAG completeness do not strongly correlate with percent of contigs binned, soil depth, or site ($R^2 < 0.2$, Table S2), suggesting that the ability to retrieve MAGs was similar in all samples. MAG assembly size increases slightly with completeness ($R^2 = 0.61$, $p < 10^{-16}$, Figure S2, Table S3). The percentage of each metagenome assembly that cumulatively maps to any of the

169 MAGs varies from 8.3% (BPF1 56-58 cm) to 87.89% (BPF4 0-2 cm) (Figure S3, Tables S2, S3).

MAGs belong to 12 phyla. *Acidobacteriota* account for the most MAGs (n = 60) from 5 of this phylum's 13 classes (*Blastocatellia* 37 MAGs, *Vicinamibacteria* 13, *Thermoanaerobaculia* 8, *Acidobacteriae*, and *Mor1* 1). *Actinomycetota* have the second highest number of MAGs (n = 52) spanning four of this phylum's nine classes (*Thermoleophilia* 45 MAGs, *Acidimicrobiia* 3, *Actinomycetia* 2, and *UBA4738* 2). *Chloroflexota* have 24 MAGs from three classes (*Ellin6529* had 21 MAGs, *Anaerolineae* 2, and *Dehalococcoidia* 1). *Proteobacteria* have 16 MAGs (5 from *Alphaproteobacteria* and 11 from *Gammaproteobacteria* classes). *Gemmatimonadota* have 3 MAGs all from class *Gemmatimonadetes*. *Planctomycetota* have 5 MAGs, 3 in *Phycisphaerae* and 2 in *Planctomycetes* classes. Three MAGs are assigned to *Verrucomicrobiota*, all within the class *Verrucomicrobiae*. Four phyla are represented by one MAG each: *Bacteroidota*, *Cyanobacteria*, *Eisenbacteria*, and *Methlomirabilota*. The only archaeal MAGs are 2 *Thermoproteota*, both in the class *Nitrososphaeria*. These MAGs represent the most abundant phyla as shown in the (unbinned) metagenomic assemblies (Table S4).

MAG read mapping to all metagenomic samples

The three most abundant phyla by metagenomic read-mapping across all samples are *Acidobacteriota*, *Actinomycetota*, and *Chloroflexota* (Figures 3, S4, and S5). Thirty of the most abundant MAGs from the major clusters discussed below are shown in Figure 3 because the complete heatmap with 169 MAGs was too large to be easily visualized in the main manuscript (Fig. S5). The *Acidobacteriota* classes *Thermoanaerobaculia* and *Blastocatellia* vary greatly in abundance with depth in the upper 20 cm. *Actinomycetota* (mostly the class *Thermoleophilia*) and *Chloroflexota* (mostly the class *Ellin6529*) are consistently abundant regardless of depth below the upper few centimeters (Figures 4 and S3), suggesting that any possible depth-related abundance trends for these groups may occur at the lower family or genus levels.

The highest read abundance for each MAG is from the assembly from which they originated, followed by assemblies from the same site (Figure 3 and S5). Ninety-two MAGs recruited reads from other site assemblies while 57 MAGs are constrained to a subset of locations or depths. Only 20 MAGs (9 *Actinomycetota*, 3 *Chloroflexota*, 1 *Proteobacteria*, 1 *Gemmatimonadota*, 2 *Planctomycetota*, 1 *Eisenbacteria* and 3 *Acidobacteriota*) recruited reads from all 56 samples (Figure 3, S3, S5). A Spearman correlation-based matrix between all the MAG relative abundances at the sites was used to determine relationships between MAGs, based on their read abundances across the 56 metagenome samples. Stronger correlation coefficients (higher than 0.5) between MAGs result in closer dendrogram clustering (Figure S5). Dendrogram nodes do not always cluster by phylum or MAG's origin sample assembly, but rather by abundance similarities; for example, *Actinomycetota* do not all cluster in the same dendrogram branch nor do all BPF1 MAGs cluster together. This suggests that the approach of recruiting reads from all metagenomes to all MAGs assembled from the different soil samples is an effective way to explore the ecological significance of the MAGs. Reads from most MAGs are present in at least two sites, with a few MAGs from the same phylum found only in a single borehole site. For instance 11 *Actinomycetota* *Blastocatellia* MAGs are only in BPF2; 12 *Acidobacteriota* *Vicinamibacteria* MAGs are only below 22 cm in BPF1; 8 *Actinomycetota* *Blastocatellia* MAGs are only in the upper 32 cm of BPF1, 5 *Proteobacteriota* *Gammaproteobacteria* MAGs are found at 24-40 cm and 50-58 cm in BPF1; 3 *Planctomycetota* *phycisphaerae* MAGs are in low abundance and only present above 50 cm in BPF1 (Figure 3).

Dendrogram clustering based on Spearman rank correlations identified three major patterns in groupings of MAGs (Figure S5): A) those that increase abundance with depth but were predominantly found in BPF1; B) those that increase abundance with depth in both winter and summer cores but are more abundant in winter core BPF2; and C) those that decrease abundance with depth in all cores. The most abundant MAGs from each of these three clusters are shown in Figure 3. The microbial communities vary with depth, and microbial abundances persist between the frozen cores (BPF1 and BPF2) vs the thawed soils (BPF3, BPF4, and BPF5). The phylum level membership is similar in these three clusters (Figure 3 and S6). At the class level, however, *Acidobacteriota* differentiate within the three clusters. In cluster C, *Thermoanaerobaculia* Gp7-AA8 is in both frozen cores (5 MAGs are present in all 6 samples from the upper 12 cm of BPF1) and thawed cores (2 MAGs are abundant in all 3 thawed cores). Farther (clusters A and B), *Blastocatellia* UBA7656 MAGs are dominantly found in the frozen cores 6-28 cm (12 MAGs) and 40-48 cm (4 MAGs) in BPF1 and 0-30 cm in BPF2 (15 MAGs), with the 12 BPF1 MAGs and 13 BPF2 MAGs recruiting reads only from their own core. 92% of *Blastocatellia* UBA7656 MAGs are present only in frozen cores. *Acidobacteriota* order *Vicinamibacterales* increases in abundance in frozen core BPF1 below 20 cm (19 MAGs in layer 20-58 cm) and is absent in all thawed cores. *Pyrinomonadales* is present in similar abundance in all samples, though different MAGs recruited reads from the frozen (8 MAGs) and thawed (3 MAGs) samples.

Actinomycetota are more broadly distributed along the depth gradient and among sites than other phyla, with each class recruiting reads from all depths of winter cores and summer samples. However, some family- and genus-level taxonomic groups of *Actinomycetota* are present only at a subset of depths. Thus, the family *Thermoleophilaceae* within the class *Solirubrobacterales* are abundant below 24 cm in all cores. In contrast, the members of the family *Gaiellaceae* that could not be classified at the genus level are more abundant in the upper 30 cm in all cores. The *Actinomycetota* lineages such as order *Solirubrobacterales* family 70-9, family *Gaiellaceae* genus VAXT01 and family *Gaiellaceae* unclassified contain some MAGs that are mostly found at deeper layers and others that are more abundant at upper depths.

Almost all *Chloroflexota* MAGs increase in abundance with depth (17 MAGs in winter cores and 3 MAGs in both winter and summer samples); only 3 MAGs of *Chloroflexota* Ellin6529 originating from the summer samples are abundant in the surface. *Alphaproteobacteria* MAGs are more often found at the surface, while *Gammaproteobacteria* MAGs are present at all depths in both winter and summer samples except only below 22 cm in winter core BPF1. *Planctomycetota* MAGs are usually more often found at the surface. *Gemmatimonadota* MAGs are present at all depths in both winter and summer samples. *Verrucomicrobiota* and *Nitrosphaeria* MAGs are either present in both summer and surficial winter cores or found in deeper layers in winter cores and are not found in the summer samples. The *Methylomirabilota* MAG is abundant at deeper layers in frozen cores and is not found in summer samples. *Eisenbacteria* MAGs are in deeper samples in both winter and summer samples. The single *Cyanobacteria* MAG is present in the surface sample of summer core BPF5. The *Bacteroidota* MAG is present in the upper 4 cm of winter core BPF1.

The fraction of the cumulative total of reads mapped from each assembly to any of the 169 MAGs decrease with increasing sample depth (Figure S2). The surface samples show the highest cumulative fraction of reads mapped to the residing MAGs for those samples, while the low fraction of reads mapped from the deeper samples suggests that the community members inhabiting these layers are not a complete assessment of microorganisms that are present in the permafrost active layer. To obtain substantial diversity coverage, a higher metagenomic

sequencing depth with more source material is required. Additionally, the higher fraction of reads mapped in surface assemblies suggest that those MAGs are a better representation of the microbes present in the surface samples.

Correlation networks

Co-existence networks evaluated using monotonic Spearman's correlation with coefficient of ≥ 0.7 reveal that distributions of the 169 MAGs correlate to those of other MAGs from the same location or depth but not from different sites (Figure 4 A). At some depths, MAGs strongly correlate with each other (orange to red color, Figure 4 A), but MAGs from the same depths do not always have strong correlations. Layers between 16 and 48 cm in BPF1 contain 58 MAGs, which form a large network where each MAG correlates with a minimum of two others. Thus, two *Acidobacteriota* MAGs in this cluster each correlate with 10 other MAGs. BPF2 show similar results, with a tighter network in the middle of the core with 14 MAGs that correlate strongly, while the clusters above and below it do not correlate with MAGs from other depths. All 10 MAGs in BPF3 correlate to each other at a maximum of four connections. Limited to selected layers, BPF4 shows three groups of correlating MAGs at different depths with the highest Spearman's correlation coefficients at 43-45 cm (Figure 4 A). In BPF5, no MAGs were available from deeper layers, while all 6 MAGs from upper layers correlated through the three *Actinomycetota* present, one at each of the sample depths.

At the phylum level, MAGs form two complete correlated groups, suggesting that similar environmental factors may drive abundances of these phyla. Multilateral correlation occurs between *Actinomycetota*, *Chloroflexota*, and *Proteobacteria* (group 1) and *Actinomycetota*, *Chloroflexota*, *Methylomirabilota* and *Proteobacteria* (group 2) (Figure 4 B). Two groups with incomplete correlations were also observed. Thus, *Acidobacteriota* correlate with *Thermoproteota* and *Gemmatimonadota*, but the latter two phyla do not correlate with each other. *Plantomycetota* correlate positively with both *Verrucomicrobiota* and *Bacteroidota*, but the latter two phyla do not correlate with each other.

Autotrophic and heterotrophic genes

The only carbon fixation pathway identified in the MAGs is the Calvin-Benson Cycle. It is present in 10 MAGs of *Actinomycetota* (classes *Thermoleophilia* and *UBA4738*), 2 *Chloroflexota* (class *Ellin6529*), and 1 *Cyanobacteria* (Figures 3, S5, S6). Each of 169 MAGs contain at least one gene encoding one of the seven extracellular carbon-degrading enzymes (Figure 5, Tables S5, S6), which were previously assayed in these cores and in the isolates derived from these cores [56]. Leucine aminopeptidase and phosphatase have the highest gene counts per MAG (Figure 5, Table S6), while α -glucosidase and β -D-cellubiosidase have the lowest gene counts per MAG. In BPF1, the only core spanning 60 cm, β -glucosidase, N-Acetyl- β -D-glucosaminidase, and β -xylosidase gene counts per MAG increase below ~30 cm.

Metaproteomics of BPF2

Peptides from the single metaproteomics library (BPF2 15-30 cm) which were mapped to the MAGs were classified into five phyla, with at least one representative from each of 26 COG categories, yielding 793 proteins (Figure S7 A). The *Acidobacteriota* are associated with 84% of the metaproteome ($n = 669$). Most of these ($n = 286$) are from class *Blastocatellia*, with smaller contributions from *Mor1* and *Thermoanaerobaculia* (Figure S7 B). The next two COG categories contain 280 and 107 proteins, however, those proteins are 'unidentifiable as a COG category' and

‘Function Unknown’, respectively. The COG category ‘Post-translational modifications, protein turnover and chaperons’ has 88 identified proteins, mainly from *Acidobacteriota* (79/88) (Figure S7 B).

MAGs for SeqCode

Only 43 MAGs met the requirements of $\geq 90\%$ completeness and $\leq 5\%$ contamination, and they are from *Acidobacteriota* (17 MAGs), *Actinomycetota* (15 MAGs), *Chloroflexota* (5 MAGs), *Gammaproteobacteria* (5 MAGs), and *Planctomycetota* (1 MAG). These MAGs were taxonomically classified using GTDB-Tk v2.3.2. Then MAGs were compared based on the whole-genome Average Nucleotide Identity (ANI) using FastANI v.0.1.3 and MAGs were considered to be the same species if they shared $\geq 95\%$ ANI. MAGs that were candidates to be named were selected from the same species group based on: lowest number of contigs (from 15 to 323), length of 16S rRNA from 906 bp to 1413 bp, total number of tRNA from 78 to 157, and originating from the different soil locations. A total of 19 high quality MAGs, namely *Acidobacteriota* (7 MAGs), *Actinomycetota* (6 MAGs), *Chloroflexota* (3 MAGs), *Gammaproteobacteria* (2 MAGs), and *Planctomycetota* (1 MAG) were uploaded to the NCBI under BioProject ID number PRJNA1040778. Based on MAGs’ annotation with PROKKA v1.14.6 [53], 3 *Acidobacteriota* and 1 *Actinomycetota* MAGs passed the requirements of the SeqCode such as $>75\%$ 16S rRNA gene completeness and number of contigs less than 100, so these were given names, following the nomenclature rules of SeqCode.

Two MAGs, BPF1-18-20_bin_3 and BPF2-00-02_bin_2, were taxonomically classified to be members of the class Blastocatellia, order UBA7656, family UBA7656, and genus JADGNW01 but the fact that the MAGs have 75.6% ANI between each other suggests that the MAGs belong to different species. Further analysis showed that these MAGs have $< 70\%$ ANI to 2 available Blastocatellia genomes (*Pyrinomonas methylaliphatogenes* and *Chloracidobacterium thermophilum*). MAG BPF1-18-20_bin_3 had 74.8% ANI to the species *Luteitalea pratensis* (class Vicinamibacteria), and MAG BPF2-00-02_bin_2 had 77.4% to the *Terriglobus saanensis* (class Terriglobia). There are metabolic similarities between these three classes. Acidobacteria of the classes Blastocatellia, Vicinamibacteria and Terriglobia are aerobic, psychrotolerant, grow on different simple sugars and dominate soils with a neutral or slightly basic pH in case of Blastocatellia and Vicinamibacteria [33], or at a neutral to slightly acidic pH in case of Terriglobia [33]. The fact that the MAGs show low similarity to known species support our conclusion that these MAGs belong to a new genus, family and order.

Description of high-quality MAGs submitted to SeqCode are as follows, using GTDB-Tk version 2.3.2 for classification and NCBI PGAP v.6.7 for reannotation:

MAG: BPF1-18-20_bin_3

SAMPLE: core BPF1, depth of 18-20 cm

BIOPROJECT: PRJNA1040778

MAG BIOSAMPLE: SAMN38265135

PHYSICAL BIOSAMPLE: SAMN39337272

GENBANK ASSEMBLY ACCESSION: GCA_038040485.1

GENOME ASSEMBLY: ASM3804048v1

MOLECULE TYPE: genomic DNA

PLATFORM: The sequencing platform was Illumina NovaSeq 6000 S4 (2 x 150 bp).

348 COVERAGE: The estimated depth of sequencing coverage is 2.5
349 PROGRAM: The metagenome was assembled with metaSPAdes v3.11.1 and binned using
350 MaxBin2 v2.2.5 and MetaBAT2 v2.12.1. Annotated with PROKKA v1.14.5 and reannotated with
351 NCBI PGAP v.6.7. Taxonomic classifications to bacterial genomes was assigned using GTDB-Tk
352 v2.3.2. Functional description was done with DRAM v.0.1.2.
353 ASSEMBLY NAME: MAG BPF1-18-20_bin_3
354 ASSEMBLY TYPE: MAG
355 LINEAGE: Acidobacteriota; Blastocatellia; UBA7656; UBA7656; JADGNW01
356 PROPOSED: Acidobacteriota; Blastocatellia; Order = Onstotales, Family = Onstottaceae, Genus
357 = *Onstottus*, Species = *arcticum*
358 DESCRIPTION: MAG BPF1-18-20_bin_3 has 98.01% completeness and 3.42% contamination.
359 Total length is 6,906,781 bp in 72 scaffolds; scaffold N50 is 162.4 kb; scaffold L50 is 13; the
360 number of contigs is 103; contig N50 is 121.8 kb; contig L50 is 19; the number of genes is 5,743;
361 CDSs is 5,663; the number of protein coding genes is 5,610; GC content is 57.25%; this MAG has
362 one complete 5S rRNA; one partial 23S rRNA; 74 tRNA genes; 4 ncRNAs; 53 Pseudo genes. The
363 ANI analysis with 42 available Acidobacteriota genomes showed only genome NZ_CP015136 of
364 the species *Luteitalea pratensis* (class Vicinamibacteria) had 74.8% whereas ANI percentages with
365 other genomes were <70%. CAZY Families identified were glycoside hydrolases = 1; and glycosyl
366 transferases = 26. Identified signature for conversion of nitric oxide to nitrous oxide.
367 *Onstottus arcticum* likely has an aerobic heterotrophic lifestyle, with genes encoding glycolysis
368 and tricarboxylic acid cycle, as well as a low affinity cytochrome c oxidase, F-type ATPase, and
369 succinate dehydrogenase. It may produce nitrous oxide due to the presence of a nitric oxide
370 reductase. It encodes 1 glycoside hydrolases and 26 glycosyl transferases. It was retrieved from
371 18-20 cm soil depth, sampled from site BPF1, Bayelva Permafrost site 1, while frozen in April
372 2018. Total carbon was 2.5 %, dissolved organic carbon was 0.93 mg/g soil, total nitrogen was
373 0.17 %, nitrate was 0.007 µg/g soil, and the soil C/N ratio was 14.3.
374
375
376 MAG: BPF2-00-02_bin_2
377 SAMPLE: core BPF2, depth of 0-2 cm
378 BIOPROJECT: PRJNA1040778
379 MAG BIOSAMPLE: SAMN38265137
380 PHYSICAL BIOSAMPLE: SAMN39337894
381 GENBANK ASSEMBLY ACCESSION: GCA_038040445.1
382 GENOME ASSEMBLY: ASM3804044v1
383 MOLECULETYPE: genomic DNA
384 PLATFORM: The sequencing platform was Illumina NovaSeq 6000 S4 (2 x 150 bp).
385 COVERAGE: The estimated depth of sequencing coverage is 2.5
386 PROGRAM: The metagenome was assembled with metaSPAdes v3.11.1 and binned using
387 MaxBin2 v2.2.5 and MetaBAT2 v2.12.1. Annotated with PROKKA v1.14.5 and reannotated with
388 NCBI PGAP v.6.7. Taxonomic classifications to bacterial genomes was assigned using GTDB-Tk
389 v2.3.2. Functional description was done with DRAM v.0.1.2.
390 ASSEMBLYNAME: MAG BPF2-00-02_bin_2
391 ASSEMBLY_TYPE: MAG
392 LINEAGE: Acidobacteriota; Blastocatellia; UBA7656; UBA7656; JADGNW01

PROPOSED: Acidobacteriota; Blastocatellia; Order = Onstotales, Family = Onstottaceae, Genus = *Onstottus*, Species = *frigus*

DESCRIPTION: MAG BPF2-00-02_bin_2 has 98.29% completeness and 3.42% contamination. Total length is 6,749,686 bp in 72 scaffolds; scaffold N50 is 181.9 kb; scaffold L50 is 12; the number of contigs is 102; contig N50 is 143.4 kb; contig L50 is 15; the number of genes is 5,799; DCSs is 5,722; the number of protein coding genes is 5,668; GC content is 56.88%; this MAG has one partial 16S rRNA small subunit with a length of 341 bp; one complete 5S rRNA; one complete 23S rRNA; 71 tRNA genes; 3 ncRNAs; 54 Pseudo genes. Based on 16S rRNA gene search using the NCBI Megablastn, the closest relative is Acidobacteriota from class Terriglobia and genus *Terriglobus saanensis* strain SP1PR4 at (320/341) 94% identity. The highest ANI match with 6 available *Terriglobus* genomes including *Terriglobus saanensis* strain SP1PR4 is 77.4%. CAZY Families identified were glycoside hydrolases = 2; and glycosyl transferases = 31; polysaccharide lyases = 2. Identified signature for conversion of nitric oxide to nitrous oxide, and nitrate and nitrite turnover.

Onstottus frigus likely has an aerobic heterotrophic lifestyle, with genes encoding glycolysis and tricarboxylic acid cycle, as well as a low affinity cytochrome c oxidase, F-type ATPase, and succinate dehydrogenase. It may produce nitrous oxide due to the presence of a nitric oxide reductase, and may reduce nitrate and nitrite. It encodes 2 glycoside hydrolases and 31 glycosyl transferases. It was retrieved from 0-2 cm soil depth, sampled from site BPF2, Bayelva Permafrost site 2, while frozen in April 2018. Total carbon was 1.9 %, dissolved organic carbon was 2.55 mg/g soil, total nitrogen was 0.02 %, nitrate was 0.02 µg/g soil, and the soil C/N ratio was 45.1.

MAG: BPF1-10-12_bin_2

SAMPLE: core BPF1, depth of 10-12 cm

BIOPROJECT: PRJNA1040778

MAG BIOSAMPLE: SAMN38265139

PHYSICAL BIOSAMPLE: SAMN39338193

GENBANK ASSEMBLY ACCESSION: GCA_038040385.1

GENOME ASSEMBLY: ASM3804038v1

MOLECULE TYPE: genomic DNA

PLATFORM: The sequencing platform was Illumina NovaSeq 6000 S4 (2 x 150 bp).

COVERAGE: The estimated depth of sequencing coverage is 2.5

PROGRAM: The metagenome was assembled with metaSPAdes v3.11.1 and binned using MaxBin2 v2.2.5 and MetaBAT2 v2.12.1. Annotated with PROKKA v1.14.5 and reannotated with NCBI PGAP v.6.7. Taxonomic classifications to bacterial genomes was assigned using GTDB-Tk v2.3.2. Functional description was done with DRAM v.0.1.2.

ASSEMBLY NAME: MAG BPF1-10-12_bin_2

ASSEMBLY TYPE: MAG

LINEAGE: Acidobacteriota; Thermoanaerobaculia; Gp7-AA8; Gp7-AA8; none

PROPOSED: Acidobacteriota; Thermoanaerobaculia; Order = Gilichinskyiales, Family = Gilichinskyiaceae, Genus = *Gilichinskyi*, Species = *gelida*.

DESCRIPTION: MAG BPF1-10-12_bin_2 has 99.1% completeness and 4.56% contamination. Total length is 5,458,401 bp in 33 scaffolds; scaffold N50 is 244.3 kb; scaffold L50 is 7; the number of contigs is 47; contig N50 is 196.8 kb; contig L50 is 10; the number of genes is 5,245; CDSs is 5,182; the number of protein coding genes is 5,125; GC content is 65.9%; this MAG has

one partial 16S rRNA small subunit with length of 981 bp; one complete 5S rRNA; one partial 23S rRNA; 56 tRNA genes; 4 ncRNAs; 57 Pseudo genes. The best ANI data match is 80.18%. Search for highly similar 16S rRNA gene sequences using the NCBI Megablastn showed highest identity with *Thermoanaerobaculum aquaticum* strain MP-01 at (819/978) 84%. CAZY Families identified were auxiliary activities = 2; glycoside hydrolases = 7; and glycosyl transferases = 23. As opposed to the closest phylogenetic relative *Thermoanaerobaculum aquaticum* strain MP-01, which is anaerobic thermophile isolated from a freshwater hot spring, the proposed species *Gilichinskyi gelida* likely has an aerobic heterotrophic lifestyle, with genes encoding glycolysis and tricarboxylic acid cycle, as well as a low affinity cytochrome c oxidase, F-type ATPase, and succinate dehydrogenase. It encodes seven glycoside hydrolases and 23 glycosyl transferases. It was retrieved from 10-12 cm soil depth, sampled from site BPF1, Bayelva Permafrost site 1, while frozen in April 2018. Total carbon was 2.0 %, dissolved organic carbon was 1.17 mg/g soil, total nitrogen was 0.14 %, nitrate was 0.04 µg/g soil, and the soil C/N ratio was 14.5.

MAG: BPF1-56-58_bin_1

SAMPLE: core BPF1, depth of 56-58 cm

BIOPROJECT: PRJNA1040778

MAG BIOSAMPLE: SAMN38265143

PHYSICAL BIOSAMPLE: SAMN39398982

GENBANK ASSEMBLY ACCESSION: GCA_038040325.1

GENOME ASSEMBLY: ASM3804032v1

MOLECULE_TYPE: genomic DNA

PLATFORM: The sequencing platform was Illumina NovaSeq 6000 S4 (2 x 150 bp).

COVERAGE: The estimated depth of sequencing coverage is 2.5

PROGRAM: The metagenome was assembled with metaSPAdes v3.11.1 and binned using MaxBin2 v2.2.5 and MetaBAT2 v2.12.1. Annotated with PROKKA v1.14.5 and reannotated with NCBI PGAP v.6.7. Taxonomic classifications to bacterial genomes was assigned using GTDB-Tk v2.3.2. Functional description was done with DRAM v.0.1.2.

ASSEMBLY_NAME: MAG BPF1-56-58_bin_1

ASSEMBLY_TYPE: MAG

LINEAGE: Actinomycetota; Thermoleophilia; Solirubrobacterales; 70-9

PROPOSED: Actinomycetota; Thermoleophilia; Solirubrobacterales; Family = Mayfieldaceae, Genus = *Mayfieldus*, Species = *profundus*.

DESCRIPTION: MAG BPF1-56-58_bin_1 has 92.41% completeness and 1.29% contamination. Total length is 2,308,231 bp in 95 scaffolds; scaffold N50 is 46.1 kb; scaffold L50 is 14; the number of contigs is 99; contig N50 is 44.1 kb; contig L50 is 15; the number of genes is 2,390; CDSs is 2,344; the number of protein coding genes is 2,325; GC content is 69.56%; this MAG has one partial 16S rRNA small subunit with length 86 bp; one complete 5S rRNA; one partial 23S rRNA; 40 tRNA genes; 3 ncRNAs; 19 Pseudo genes. A search for highly similar 16S rRNA gene sequences using the NCBI Megablastn showed the closest relative in the order Solirubrobacterales is genus *Conexibacter woesei* DSM 14684 at (88/89) 99% identity. Nevertheless, the highest ANI match with 13 available genomes in the order Solirubrobacterales, including genome *C. woesei* DSM 14684, is 76.7%. The genome *C. woesei* DSM 14684 has a size of ~6.4 Mb; 5,998 total genes; and 4,466 genes with function prediction, while the strain is strictly aerobic and grows at temperatures of 28°-37°C [50]. CAZY Families identified were auxiliary activities = 1; glycoside

hydrolases = 2; and glycosyl transferases = 24. We also identified the signature for conversion of nitrite oxide to nitric oxide.

Mayfieldus profundus likely has an aerobic heterotrophic psychrotolerant lifestyle, with genes encoding glycolysis and tricarboxylic acid cycle, as well as a low affinity cytochrome c oxidase, F-type ATPase, and succinate dehydrogenase. It may reduce nitrite. It encodes 2 glycoside hydrolases and 24 glycosyl transferases. It was retrieved from 0-2 cm soil depth, sampled from site BPF2, Bayelva Permafrost site 2, while frozen in April 2018. Total carbon was 1.9 %, dissolved organic carbon was 2.55 mg/g soil, total nitrogen was 0.02 %, nitrate was 0.02 µg/g soil, and the soil C/N ratio was 45.1.

Discussion

The active layer soils of the Bayelva site near Ny Ålesund, Svalbard, experience large variations in temperature and VWC between depths and seasons/locations (Figure 2). Thawing and freezing are transmitted from the top of the soil downward seasonally (Table 1). Since the soils remain close to 0°C much of the year, the vertical temperature transmission is buffered by phase changes between ice and VWC. Microorganisms in deeper layers have longer exposures to VWC in the fall and upper layers have earlier access to liquid water in the spring. The presence of VWC and its association with osmolarity can significantly impact the microbial community [52,72]. The entire seasonal temperature fluctuations at 40 cm depth span 20°C, notably from -10°C to 10°C between April and September [5]. Other geochemical parameters, such as total inorganic and organic carbon, total nitrogen, and the stable isotopic ratios of carbon and nitrogen, vary gradually with depth as shown previously for BPF1 and BPF2 [56].

Previous 16S rRNA gene amplicon surveys showed systematic changes at the phylum level with depth of the active layer soils in Svalbard [39] and Alaska [64], as well as seasonal variations in the active layer bacterial community due to dramatic annual changes in temperature and soil chemistry [52], suggesting that VWC and osmolarity might serve as key factors driving the microbial community. Our approach focused on MAG read abundance, since this avoids the bias of primer amplification (16S rRNA gene amplicon surveys omit organisms whose 16S rRNA genes do not match the primers) and allows a more precise taxonomic resolution. However, MAG assembly and binning has the caveat that rare microorganisms from phyla with high intraphylum diversity may be missed [15]. The metaproteomics and MAGs' analysis corroborate that *Acidobacteriota*, *Actinomycetota*, and *Chloroflexota* are the most abundant phyla found in permafrost active layer samples [11,31,40,64,71]. The fact that most of the MAGs are from uncultured genera, or even families, classes, or orders, emphasizes the vast unknown diversity within this environment.

Phylum *Acidobacteriota*

60 MAGs are from *Acidobacteriota*, more than from any other phylum in this data set. *Adicobacteriota* is one of the most abundant and diverse phyla on Earth (28) that comprises up to 52% of the soil microbial community commonly found in surficial soils [60,75]. *Acidobacteriota* were found to dominate a thaw gradient in northern Sweden from carbon-rich palsa through a bog (thawing permafrost) to a fen (no underlying permafrost) [37] and during increases in soil respiration in a range of Antarctic soils [76], suggesting they are particularly important as soils warm.

The *Acidobacteriota Thermoanaerobaculia* Gp7-AA8 MAGs are dominant in the surface layer, suggesting a high tolerance to freeze-thaw and desiccation in surficial soils. Other studies

have indicated metabolic versatility of members of the class *Thermoanaerobaculia* that were found to have pathways for sulfate reduction in low-temperature marine sediments of Svalbard [14], and genes for reductive carboxylation that were identified in an isolate from a hot spring with the ability to grow as strict anaerobic chemoorganotroph on proteinaceous compounds [59]. Acidobacterial class *Blastocatellia* is represented by two families, UBA7656 and *Pyrinomonadales*. The family UBA7656 is found only below 6 cm, while dominating both the metagenome and metaproteome described in this study. MAGs from the family *Pyrinomonadales* are ubiquitous. Based on OTU analyses, the class *Blastocatellia* was shown to be one of the most common in tundra aeolian sands and was associated with unvegetated habitats (5). The presence of MAGs from the acidobacterial class *Vicinamibacteria* increase below 20 cm. A recent metabolic reconstruction study of *Acidobacteriota* MAGs from activated sludge wastewater treatment plants showed that *Acidobacteriota* including *Vicinamibacteria* are involved in nitrogen and phosphorus removal and iron reduction and utilization of organic compounds like xylose, acetate and fatty acids [27].

Collectively, the phylum *Acidobacteriota* shows a great deal of metabolic flexibility. The vertical depth variations in abundance of different acidobacterial classes is supported by properties and characteristics that are inherent to *Acidobacteriota*, such as natural resistance to multiple cycles of freezing and thawing [34], tolerance to drought conditions [7,78], and ability to withstand variable soil conditions including nutrient fluctuations and water content variations (35). Therefore, our study suggests that *Acidobacteriota* are better equipped to respond quickly to climate change than other clades and their higher relative abundance may occur due to recent rapid changes caused by ongoing warming.

Two lineages of *Acidobacteriota*, the *Vicinamibacteria* *Vicinamibacterales* and *Blastocatellia* UBA7656, recruit almost no reads from the summer samples. Either we missed the depths in the thawed soil that contained these clades, or these microorganisms develop better at low-temperature conditions during winter season (November-April) and become less dominant when the season shifts and the active layer thaws. Another possibility is that this is accounted for by the high degree of spatial variability in microbial communities previously noted at this site [57]. The possibility that we missed the samples containing these clades is supported by results from a recent study showing that *Acidobacterial* families, e.g., *Vicinamibacteraceae* and *Blastocatellaceae*, are equally abundant in 16S rRNA gene amplicon libraries in surficial soils collected during summer and winter near Ny Ålesund [30]. However, we cannot exclude the possibility that some members of the microbial population could be adapted to negative temperature and high osmolarity conditions, as has been observed previously in activemicrobial populations [12] that survive due to presence thin films of unfrozen brine [16] in frozen soils.

Phyla *Actinomycetota* and *Chloroflexota*

Actinomycetota and *Chloroflexota* tend to increase in relative abundance with depth and do not form subgroups that occupy particular depths, like the *Acidobacteriota* do. These phyla are found in a variety of Arctic environments, including ancient Siberian permafrost [55], old Canadian permafrost [17], Alaskan cave permafrost [8], active layer and upper permafrost of the Canadian High Arctic [73], and thermokarst bog [71]. Characteristically for these phyla, the same MAGs often recruited reads from all cores, which were distributed spatially and seasonally, with similar depth trends in all of them. This suggests that similar populations of *Actinomycetota* and *Chloroflexota* are widespread spatially and persist year-round, with depth trends driven by their tolerance to the depth-related changes in exposure to VWC. Similar to the current study, previous

analyses of 16S rRNA gene distributions showed that many types of *Actinobacteria* and *Chloroflexota* were present in both thawed and frozen samples from surficial soils around Ny Ålesund, Svalbard [39] and another study reported an increase in relative abundance of *Actinomycetota* and *Chloroflexota* with depth of permafrost soil [30]. Our results suggest that this pattern is true for all classes within the *Actinobacteria* and *Chloroflexota*. The fact that *Actinomycetota* are more evenly distributed with depth and more abundant in deeper layers than *Acidobacteriota* is consistent with other active layer and permafrost research [28,75].

Actinomycetota have the second most individual MAGs, the second most abundant proteins, and the most cumulative reads in every core. The most abundant class of *Actinomycetota*, *Thermoleophilia*, was shown to be abundant in soil and hot springs and contain genes for carbon-nitrogen hydrolase (42). The *Actinomycetota* class UBA4738 is a relatively new classification and MAGs of this class derived from the Antarctic Mackay Glacier soils were shown to be able to use atmospheric H₂ and CO to meet energy, carbon, and hydration needs [43]. MAGs from the *Thermoleophilia* and UBA4738 classes contain the Calvin-Benson-Bassham cycle (CBB pathway) (Figure S6). While *Actinomycetota* have not been shown to be widely autotrophic, some MAGs previously characterized contain autotrophic pathways [54], including other carbon fixation pathways, like Wood-Ljungdahl [36]. Two *Chloroflexota* MAGs also contain the CBB pathway. MAGs containing autotrophic pathways were found in all samples, demonstrating the potential for carbon fixation at all depths. The only MAG of *Actinomycetota* class *Acidimicrobiia* is highly abundant in all soil samples studied here, possibly due to its putative ability to contribute to iron cycling as suggested for a *Acidimicrobiia* MAG discovered in thawed discontinuous permafrost in Sweden (32). Cultured isolates of *Actinomycetota* class *Actinomycetia* retrieved from sulfide ore deposits in Russia were shown to consume O₂ and emit CO₂ (41). Given the abundance of *Actinomycetota* in the Bayelva active layer, they might play a central role in carbon cycling there.

The correlation networks suggest that phyla *Actinomycetota* and *Acidobacteriota* inhabit different niches (Figure 4B). The phylum *Chloroflexota* co-correlated with three other phyla (*Actinomycetota*, *Proteobacteria*, and *Methylomirabilota* with Spearman coefficient of 0.9, 0.7, and 0.7; respectively, Figure 4B). These phyla were previously found in cold Arctic habitats. Thus, *Chloroflexota*, *Actinomycetota*, and *Proteobacteria* were repeatedly found in active layer soils in the high Arctic of Canada and Sweden [32,69], whereas *Methylomirabilota* was discovered in the bottom water of permafrost thaw pond in the Northern Canada (39). Positive correlation between *Actinomycetota* and *Chloroflexota* was shown in permafrost of the Greater Khingan Mountain region where they increased in abundance with depth [28]. These two phyla may correlate with each other due to their ability to adapt to low nutrient environments and use recalcitrant organic compounds [26].

Other studies of the active layer/permafrost cores showed a shift from *Acidobacteriota* and *Proteobacteria*-dominated microbial community in the active layer to *Actinomycetota*, *Bacteroidetes*, *Chloroflexota* and *Proteobacteria*-dominated community in the permafrost up to 2 m down (16). *Actinomycetota* are able to survive in more oligotrophic conditions than *Acidobacteriota* which prefer copiotrophic environments [28]. The episodic distribution of *Acidobacteriota* in our study suggests that they evolve quickly in response to availability of organic matter substrates due to changes from seasons or spatial heterogeneity, whereas *Actinomycetota* and *Chloroflexota* are less responsive to small-scale changes. Likewise, *Acidobacteriota* classes' vertical distribution may follow patterns according to their desiccation tolerance, with *Thermoanaerobaculia* Gp7-AA8 being the most tolerant, followed by *Blastocatellia* UBA7656 and then *Vicinamibacteria* *Vicinamibacterales*. Distribution and

abundance of *Actinomycetota* and *Chloroflexota* correlated positively with the increase of water content with depth.

Carbon-degrading potential, as shown by organic carbon hydrolyzing enzymes

Five of the seven organic carbon-degrading enzymes assayed previously [56] were abundant at all depths of the cores (Figure 5). The putative ability of bacteria to excrete these enzymes was demonstrated through metaproteomic discovery of proteins from COG categories for export/secretion in five phyla (Figure S7A). Genes encoding α -glucosidase and cellobiosidase were in low abundance (Figure 5), commensurate with the low rates of potential activity for these enzymes measured previously [56]. In general, predicted enzymatic activity based on the MAGs' count of enzyme-encoding genes parallel the results of enzymatic activity measured from the bulk soil samples and cultured isolates. Thus, gene counts for leucine aminopeptidase are the highest, followed by phosphatase; and both of these decrease with depth, matching the results from enzyme assays [56]. Gene counts for three enzymes (β -glucosidase, N-acetyl- β -D-glucosaminidase, and xylosidase) increase in abundance with depth, suggesting that primary degradation of soil organic matter occurs below the surface at depths where environmental conditions are more stable. The latter statement is corroborated by the results that these three enzymes likely are psychrophilic with higher catalytic activity at 5°C and 15°C than at 25°C [56]. Two of these (β -glucosidase and N-acetyl- β -D-glucosaminidase) were shown to have similar potential activities in surface and deeper samples [56]. Two-year-long monitoring effort during this study demonstrated that soil temperature is more vertically homogenous than the VWC. The VWC and the duration water stays liquid increases with depth in the upper 80 cm of active layer soils near Ny Ålesund, Svalbard (Figure 2). Therefore, these results demonstrate that more liquid water and less frequent freeze-thaw events deeper in the soil profile may support development of microbial communities with higher capacity for organic matter degradation as opposite to the surface layer (Figure 2, 6).

Conclusion

We discovered a highly heterogenous microbial community that occupies different depth-dependent soil niches showing higher organic matter degrading capabilities in deeper horizons of permafrost-affected active layer. Microbes in permafrost-affected soils are likely to experience a dramatic environmental shift as the climate continues to warm [5] leading to permafrost degradation, ground ice thawing, and changes in soil hydrology. Our results suggest that different classes of *Acidobacteriota* may react differently to these vertically varying environmental conditions, with some classes specialized for the longer periods with low VWC in surficial layers. *Actinomycetota* and *Chloroflexota*, on the other hand, are more broadly distributed and active, perhaps driving organic matter degradation in deeper layers at the forefront of permafrost thaw, which may have greater exposure to liquid water due to insulation from upper layers.

Methods

Sample collection

Two active layer soil cores, BPF1 (58 cm) and BPF2 (30 cm), were collected in April 2018 from the Bayelva site, Ny Ålesund, Svalbard, (78°55.237'N, 011°50.495'E) 84 m apart from each other, as described previously [56]. We will call these the winter core samples, since they were taken at the end of winter when the ground was still frozen. Additionally, small soil pits were dug from three other sites in early September 2019 at three additional sites in the same area: BPF3 (1-3 cm and 18-20 cm maximum depth), BPF4 (1-3 cm and 43-45 cm) and BPF5 (1-3 cm and 43-45

cm; Figure 1). We will call these the summer samples since they were taken at the end of summer before the soils re-froze for the winter. These are on Bayelva hill and are not directly influenced by seasonal glacial melt water flow which runs in the large (flood)plain channel. BPF1 and BPF2 cores were sliced in 2 cm increments with a sterile knife using aseptic techniques in the Kings Bay Marine Laboratory in Ny Ålesund, Svalbard. For BPF3, BPF4, and BPF5, four depths were cut from the wall of soil pits using a sterile scraper, put directly into sterile Whirl-Pak® sample bags, and frozen within hours. Samples were transported frozen to the laboratory at the University of Tennessee in Knoxville where they were stored at -80°C until analysis.

Permafrost thermal and hydrologic properties

We investigated the thermal and hydrologic properties of the Bayelva long-term permafrost monitoring site using a quality-controlled dataset of soil temperature and VWC [5]. We analyzed the thermal and hydrologic properties during four seasons of 2018 and 2019 (Figure 2), which overlaps with sampling events for winter cores BPF1 and BPF2 as well as for summer pit samples BPF3, BPF4, and BPF5. The start of fall re-freeze was defined as when soil temperatures dropped and remained below -0.1°C. The period of zero curtain was defined as a constant temperature around 0°C until the hourly soil temperature dropped and remained below -0.5°C. The mean VWC for the last day of the zero-curtain period was calculated over a 24-hour period. Methods are more extensively described in [5], and all data are available online (<https://doi.pangaea.de/10.1594/PANGAEA.882060>).

DNA extraction, sequencing, and bioinformatic analysis

DNA was extracted from ~0.5 g taken from the inside middle portion of soil sample and extracted with a QIAGEN PowerSoil DNA Kit (Hilden, Germany) within 24 h of collection (BPF1 and BPF2) or after being transported frozen to the University of Tennessee Knoxville (BPF3, BPF4, and BPF5). DNA libraries were prepared using the SMARTer ThruPlex DNA-seq libraries (350 bp average fragment size) and sequenced on Illumina NovaSeq 6000 S4 (2 x 150 bp) at Science for Life Laboratory in Stockholm, Sweden.

All bioinformatic operations were performed using the Infrastructure for Science Applications and Advanced Computing (ISAAC) high-performance computing resource at Oak Ridge National Laboratory, Knoxville, Tennessee USA in conjunction with the University of Tennessee, Knoxville, Tennessee, USA. Each of 56 libraries were assembled and binned separately. MetaWrap v1.3.2 [68] was used for read quality control, assembly was done with metaSPAdes v3.11.1 [42], and binning was performed using MaxBin2 v2.2.5 [74] and MetaBAT2 v2.12.1 [24]. Bins were refined on a medium, or higher, quality category [6] of ≥50% completeness and <10% contamination scores via CheckM v1.4.0 [45] and taxonomically identified with GTDB-Tk v1.5.1 [44]. Taxonomic identities of unbinned metagenomic assemblies was determined with Kaiju [35]. MAG distribution was computed with metaWRAP's quant_bins module which assesses the average length-weighted abundance of each MAG in each sample via read-mapping with salmon v1.10.1 [46].

R Statistical Software v4.1.0 [70] was used to compute a Spearman's rank correlation coefficient matrix of the relative abundance of MAGs in the 56 samples. The correlation coefficients were used as connections in snetwork analyses across cores, sample depths, and phyla. MAG genes were annotated with PROKKA v1.14.6 [53] and assessed for genes encoding enzymes whose activities were assayed in two cores previously [56]: α-glucosidase (AG), β-glucosidase (BG), cellobiosidase (CB), leucine aminopeptidase (LAP), N-acetyl-b-D-glucosaminidase (NAG),

phosphatase (PHOS), and xylosidase (XYL). Genes for autotrophic pathways were identified as described previously [51] and a carbon fixation pathway was considered present if it was >60% complete and contained the pathway's key genes [3,49]. FastANI v.0.1.3 was used to compare MAGs to each other.

Metaproteomics

Metaproteomics analyses were done using 15 g of material from BPF2 only. Briefly, the soil was resuspended in a 4% SDS/0.1M NaOH lysis buffer prior to vortexing and centrifugation. The resulting supernatant was subjected to sonication to lyse bacterial cells. Protein clean-up and digestion were conducted via a protein aggregation capture (PAC) method [2] with slight modifications. An acidified aqueous (1% Formic acid) wash was included to remove soluble fulvic acids from the sample prior to proteolytic digestion. Tryptic peptides were measured by 1D RP LC-MS/MS on a Q Exactive Plus as previously described [47]. De novo-assisted database searches were conducted using PEAKS StudioX [63] against custom protein databases constructed using core and depth-matched metagenomes. Peptides with a false discovery rate of <1% and that uniquely mapped to a single phylum were used for protein inference and quantification. Functional annotations and orthology assignments for all identified proteins were found using eggNOG-mapper v2 [9].

Acknowledgments:

The authors want to thank the field teams in 2018 and 2019 that helped with this work: Lauren Mullen, Alexander B. O. Michaud and the AWIPEV base staff in Ny Ålesund. This work is funded by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020369 to KL, TV, RH, and AS) and the Simons Foundation (404586 to KL). Michael Tamino assisted with in-house python scripts. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

Metagenome data availability:

The 56 metagenomic libraries can be found under NCBI Metagenomic BioProject [PRJNA971579](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA971579).

Metaproteome data availability:

All raw mass spectra for the metaproteome measurement from this study have been deposited into the ProteomeXchange repository with accession numbers: ProteomeXchange-PXD041734; MassIVE- MSV000091780. The deposited data also includes the metagenome derived protein sequences used for the database search.

PANGAEA data citation: Boike, Julia; Juszak, Inge; Lange, Stephan; Chadburn, Sarah; Burke, Eleanor J; Overduin, Pier Paul; Roth, Kurt; Ippisch, Olaf; Bornemann, Niko; Stern, Lielle; Gouttevin, Isabelle; Hauber, Ernst; Westermann, Sebastian (2017): Soil data at station Bayelva

(1998-2017, level 1, version 1), link to archive. PANGAEA,
<https://doi.org/10.1594/PANGAEA.882060>,

References

- [1] Bardgett, R.D., Freeman, C., Ostle, N.J. (2008) Microbial contributions to climate change through carbon cycle feedbacks. *The ISME Journal* 2:8 2(8), 805–14, Doi: 10.1038/ismej.2008.58.
- [2] Batth, T.S., Tollenaere, M.A.X., R  ther, P., Gonzalez-Franquesa, A., Prabhakar, B.S., Bekker-Jensen, S., Deshmukh, A.S., Olsen, J. V. (2019) Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Molecular and Cellular Proteomics* 18(5), 1027–35, Doi: 10.1074/mcp.TIR118.001270.
- [3] Berg, I.A. (2011) Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Applied and Environmental Microbiology* 77(6), 1925–36, Doi: 10.1128/AEM.02473-10.
- [4] Bockheim, J.G. (2007) Importance of cryoturbation in redistributing organic carbon in permafrost-affected soils. *Soil Science Society of America Journal* 71(4), 1335–42, Doi: 10.2136/sssaj2006.0414N.
- [5] Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Paul Overduin, P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I., Hauber, E., Westermann, S. (2018) A 20-year record (1998-2017) of permafrost, active layer and meteorological conditions at a high Arctic permafrost research site (Bayelva, Spitsbergen). *Earth System Science Data* 10(1), 355–90, Doi: 10.5194/essd-10-355-2018.
- [6] Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloie-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M.,

- Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35(8), 725–31, Doi: 10.1038/nbt.3893.
- [7] Bu, X., Gu, X., Zhou, X., Zhang, M., Guo, Z., Zhang, J., Zhou, X., Chen, X., Wang, X. (2018) Extreme drought slightly decreased soil labile organic C and N contents and altered microbial community structure in a subtropical evergreen forest. *Forest Ecology and Management* 429, 18–27, Doi: 10.1016/J.FORECO.2018.06.036.
- [8] Burkert, A., Douglas, T.A., Waldrop, M.P., Mackelprang, R. (2019) Changes in the active, dead, and dormant microbial community structure across a pleistocene permafrost chronosequence. *Applied and Environmental Microbiology* 85(7), AEM.02646-18, Doi: 10.1128/AEM.02646-18.
- [9] Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J. (2021) eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 38(12), 5825–9, Doi: 10.1093/molbev/msab293.
- [10] Cohen, J., Screen, J.A., Furtado, J.C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., Jones, J. (2014) Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, 627–37, Doi: 10.1038/ngeo2234.
- [11] Dziurzynski, M., Gorecki, A., Pawlowska, J., Istel, L., Decewicz, P., Golec, P., Styczynski, M., Poszytek, K., Rokowska, A., Gorniak, D., Dziewit, L. (2022) Revealing the diversity of bacteria and fungi in the active layer of permafrost at Spitsbergen island (Arctic) – Combining classical microbiology and metabarcoding for ecological and bioprospecting exploration. *Science of The Total Environment* 856(September 2022), 159072, Doi: 10.1016/j.scitotenv.2022.159072.
- [12] Ernakovich, J.G., Wallenstein, M.D. (2015) Permafrost microbial community traits and functional diversity indicate low activity at in situ thaw temperatures. *Soil Biology and Biochemistry* 87, 78–89, Doi: 10.1016/J.SOILBIO.2015.04.009.
- [13] Fierer, N., Nemergut, D., Knight, R., Craine, J.M. (2010) Changes through time: Integrating microorganisms into the study of succession. *Research in Microbiology* 161(8), 635–42, Doi: 10.1016/j.resmic.2010.06.002.
- [14] Flieder, M., Buongiorno, J., Herbold, C.W., Hausmann, B., Rattei, T., Lloyd, K.G., Loy, A., Wasmund, K. (2021) Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. *The ISME Journal*, 1–22, Doi: 10.1038/s41396-021-00992-0.
- [15] Ghurye, J.S., Cepeda-Espinoza, V., Pop, M. (2016) Focus: Microbiome: Metagenomic Assembly: Overview, Challenges and Applications. *The Yale Journal of Biology and Medicine* 89(3), 353.
- [16] Gilichinsky, D., Rivkina, E., Shcherbakova, V., Laurinavichuis, K., Tiedje, J. (2003) Supercooled water brines within permafrost—an unknown ecological niche for microorganisms: a model for astrobiology. *Astrobiology* 3(2), 331–41, Doi: 10.1089/153110703769016424.

- [17] Goordial, J., Davila, A., Greer, C.W., Cannam, R., DiRuggiero, J., McKay, C.P., Whyte, L.G. (2017) Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper-arid polar desert. *Environmental Microbiology* 19(2), 443–58, Doi: 10.1111/1462-2920.13353.
- [18] Groenke, B., Langer, M., Nitzbon, J., Westermann, S., Gallego, G., Boike, J. (2023) Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer. *Cryosphere* 17(8), 3505–33, Doi: 10.5194/tc-17-3505-2023.
- [19] Hultman, J., Waldrop, M.P., Mackelprang, R., David, M.M., McFarland, J., Blazewicz, S.J., Harden, J., Turetsky, M.R., McGuire, A.D., Shah, M.B., VerBerkmoes, N.C., Lee, L.H., Mavrommatis, K., Jansson, J.K. (2015) Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521(7551), 208–12, Doi: 10.1038/nature14238.
- [20] Isaksen, K., Nordli, Ø., Ivanov, B., Køltzow, M.A.Ø., Aaboe, S., Gjeltén, H.M., Mezghani, A., Eastwood, S., Førland, E., Benestad, R.E., Hanssen-Bauer, I., Brækkan, R., Sviashchennikov, P., Demin, V., Revina, A., Karandasheva, T. (2022) Exceptional warming over the Barents area. *Scientific Reports* 2022 12:1 12(1), 1–18, Doi: 10.1038/s41598-022-13568-5.
- [21] Ivanova, A.A., Zhelezova, A.D., Chernov, T.I., Dedysh, S.N. (2020) Linking ecology and systematics of acidobacteria: Distinct habitat preferences of the Acidobacteriia and Blastocatellia in tundra soils. *PLoS ONE* 15(3), 1–19, Doi: 10.1371/journal.pone.0230157.
- [22] Jansson, J.K., Taş, N. (2014) The microbial ecology of permafrost. *Nature Reviews Microbiology* 12(6), 414–25, Doi: 10.1038/nrmicro3262.
- [23] Johnston, E.R., Hatt, J.K., He, Z., Wu, L., Guo, X., Luo, Y., Schuur, E.A.G., Tiedje, J.M., Zhou, J., Konstantinidis, K.T. (2019) Responses of tundra soil microbial communities to half a decade of experimental warming at two critical depths. *Proceedings of the National Academy of Sciences*, 201901307, Doi: 10.1073/PNAS.1901307116.
- [24] Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z. (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019(7), Doi: 10.7717/peerj.7359.
- [25] Keating, K., Binley, A., Bense, V., Van Dam, R.L., Christiansen, H.H. (2018) Combined geophysical measurements provide evidence for unfrozen water in permafrost in the Adventdalen Valley in Svalbard. *Geophysical Research Letters* 45(15), 7606–14, Doi: 10.1029/2017GL076508.
- [26] Kou, D., Yang, G., Li, F., Feng, X., Zhang, D., Mao, C., Zhang, Q., Peng, Y., Ji, C., Zhu, Q., Fang, Y., Liu, X., Xu-Ri., Li, S., Deng, J., Zheng, X., Fang, J., Yang, Y. (2020) Progressive nitrogen limitation across the Tibetan alpine permafrost region. *Nature Communications* 11(1), 1–9, Doi: 10.1038/s41467-020-17169-6.
- [27] Kristensen, J.M., Singleton, C., Clegg, L.A., Petriglieri, F., Nielsen, P.H. (2021) High diversity and functional potential of undescribed “Acidobacteriota” in Danish wastewater treatment plants. *Frontiers in Microbiology* 12(April), Doi: 10.3389/fmicb.2021.643950.
- [28] Li, X., Cui, Y., Ma, D., Song, D., Liu, L. (2022) Vertical distribution of bacterial community diversity in the Greater Khingan Mountain permafrost region. *Ecology and Evolution* 12(7), e9106, Doi: 10.1002/ECE3.9106.
- [29] Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., Crosby, L. (2018) Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *MSystems* 3(5), e00055-18.

- [30] Loganathachetti, D.S., Venkatachalam, S., Jabir, T., Vipindas, P. V., Krishnan, K.P. (2022) Total nitrogen influence bacterial community structure of active layer permafrost across summer and winter seasons in Ny-Ålesund, Svalbard. *World Journal of Microbiology & Biotechnology* 38(2), Doi: 10.1007/S11274-021-03210-3.
- [31] Mackelprang, R., Burkert, A., Haw, M., Mahendrarajah, T., Conaway, C.H., Douglas, T.A., Waldrop, M.P. (2017) Microbial survival strategies in ancient permafrost: insights from metagenomics. *The ISME Journal* 11(10), 2305–18, Doi: 10.1038/ismej.2017.93.
- [32] Mackelprang, R., Waldrop, M.P., DeAngelis, K.M., David, M.M., Chavarria, K.L., Blazewicz, S.J., Rubin, E.M., Jansson, J.K. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480(7377), 368–71, Doi: 10.1038/nature10576.
- [33] Männistö, M.K., Rawat, S., Starovoytov, V., Häggblom, M.M. (2011) *Terriglobus saanensis* sp. nov., an acidobacterium isolated from tundra soil. *Int J Syst Evol Microbiol* 61(8), 1823–8, Doi: 10.1099/IJS.0.026005-0/CITE/REFWORKS.
- [34] Männistö, M.K., Tiirola, M., Häggblom, M.M. (2009) Effect of freeze-thaw cycles on bacterial communities of arctic tundra soil. *Microbial Ecology* 58(3), 621–31, Doi: 10.1007/S00248-009-9516-X.
- [35] Menzel, P., Ng, K.L., Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 2016 7:1 7(1), 1–9, Doi: 10.1038/ncomms11257.
- [36] Merino, N., Kawai, M., Boyd, E.S., Colman, D.R., McGlynn, S.E., Nealson, K.H., Kurokawa, K., Hongoh, Y. (2020) Single-Cell Genomics of Novel Actinobacteria With the Wood–Ljungdahl Pathway Discovered in a Serpentinizing System. *Frontiers in Microbiology* 11, 1031, Doi: 10.3389/FMICB.2020.01031/BIBTEX.
- [37] Mondav, R., McCalley, C.K., Hodgkins, S.B., Frolking, S., Saleska, S.R., Rich, V.I., Chanton, J.P., Crill, P.M. (2017) Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. *Environmental Microbiology* 19(8), 3201–18, Doi: 10.1111/1462-2920.13809.
- [38] Morgalev, Y.N., Lushchaeva, I. V., Morgaleva, T.G., Kolesnichenko, L.G., Loiko, S. V., Krickov, I. V., Lim, A., Raudina, T. V., Volkova, I.I., Shirokova, L.S., Morgalev, S.Y., Vorobyev, S.N., Kirpotin, S.N., Pokrovsky, O.S. (2017) Bacteria primarily metabolize at the active layer/permafrost border in the peat core from a permafrost region in western Siberia. *Polar Biology* 40(8), 1645–59, Doi: 10.1007/s00300-017-2088-1.
- [39] Müller, O., Bang-Andreasen, T., White, R.A., Elberling, B., Taş, N., Kneafsey, T., Jansson, J.K., Øvreås, L. (2018) Disentangling the complexity of permafrost soil by using high resolution profiling of microbial community composition, key functions and respiration rates. *Environmental Microbiology* 20(12), 4328–42, Doi: 10.1111/1462-2920.14348.
- [40] Müller, S., Hübschmann, T., Kleinstüber, S., Vogt, C. (2012) High resolution single cell analytics to follow microbial community dynamics in anaerobic ecosystems. *Methods* 57(3), 338–49, Doi: 10.1016/J.YMETH.2012.04.001.
- [41] Nordli, Ø., Wyszynski, P., Gjeltén, H.M., Isaksen, K., Łupikasza, E., Niedźwiedz, T., Przybylak, R. (2020) Revisiting the extended Svalbard Airport monthly temperature series, and the compiled corresponding daily series 1898–2018. *Polar Research* 39, 1–15, Doi: 10.33265/POLAR.V39.3614.

- [42] Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27(5), 824–34, Doi: 10.1101/gr.213959.116.
- [43] Ortiz, M., Leung, P.M., Shelley, G., Goethem, M.W. Van., Bay, S.K., Jordaan, K., Vikram, S., Hogg, I.D., Makhalanyane, T.P., Chown, S.L., Grinter, R., Cowan, D.A., Greening, C. (2020) A genome compendium reveals diverse metabolic adaptations of Antarctic soil microorganisms. *BioRxiv*, 2020.08.06.239558, Doi: 10.1101/2020.08.06.239558.
- [44] Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarszewski, A., Chaumeil, P.A., Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36(10), 996, Doi: 10.1038/nbt.4229.
- [45] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7), 1043–55, Doi: 10.1101/gr.186072.114.
- [46] Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods* 14(4), 417, Doi: 10.1038/NMETH.4197.
- [47] Peters, S.L., Borges, A.L., Giannone, R.J., Morowitz, M.J., Banfield, J.F., Hettich, R.L. (2022) Experimental validation that human microbiome phages use alternative genetic coding. *Nature Communications* 13(1), Doi: 10.1038/s41467-022-32979-6.
- [48] Pörtner, H.O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M. (2019) The ocean and cryosphere in a changing climate. A special report of the intergovernmental panel on climate change. *Intergovernmental Panel on Climate Change*.
- [49] Probst, A.J., Castelle, C.J., Singh, A., Brown, C.T., Anantharaman, K., Sharon, I., Hug, L.A., Burstein, D., Emerson, J.B., Thomas, B.C., Banfield, J.F. (2017) Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO₂ concentrations. *Environmental Microbiology* 19(2), 459–74, Doi: 10.1111/1462-2920.13362.
- [50] Pukall, R., Lapidus, A., Glavina Del Rio, T., Copeland, A., Tice, H., Cheng, J.F., Lucas, S., Chen, F., Nolan, M., Bruce, D., Goodwin, L., Pitluck, S., Mavromatis, K., Ivanova, N., Ovchinnikova, G., Pati, A., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.J., Jeffries, C.D., Chain, P., Meincke, L., Sims, D., Brettin, T., Detter, J.C., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Kyrpides, N.C., Klenk, H.P., Hugenholtz, P. (2010) Complete genome sequence of *Conexibacter woesei* type strain (ID131577 T). *Stand Genomic Sci* 2(2), 212–9, Doi: 10.4056/SIGS.751339/TABLES/4.
- [51] Rogers, T.J., Buongiorno, J., Jessen, G.L., Schrenk, M.O., Fordyce, J.A., de Moor, J.M., Ramírez, C.J., Barry, P.H., Yücel, M., Selci, M., Cordone, A., Giovannelli, D., Lloyd, K.G. (2022) Chemolithoautotroph distributions across the subsurface of a convergent margin. *The ISME Journal* 2022 17:1 17(1), 140–50, Doi: 10.1038/s41396-022-01331-7.
- [52] Schostag, M., Stibal, M., Jacobsen, C.S., Bælum, J., Tas, N., Elberling, B., Jansson, J.K., Semenchuk, P., Priemé, A. (2015) Distinct summer and winter bacterial communities in the active layer of Svalbard permafrost revealed by DNA- and RNA-based analyses. *Frontiers in Microbiology* 6(APR), 399, Doi: 10.3389/fmicb.2015.00399.
- [53] Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14), 2068–9, Doi: 10.1093/bioinformatics/btu153.

- 987 [54] Selensky, M.J., Masterson, A.L., Blank, J.G., Lee, S.C., Osburn, M.R. (2021) Stable
988 carbon isotope depletions in lipid biomarkers suggest subsurface carbon fixation in lava
989 caves. *Journal of Geophysical Research: Biogeosciences* 126(7), e2021JG006430, Doi:
990 10.1029/2021JG006430.
- 991 [55] Sipes, K., Almatari, A., Eddie, A., Williams, D., Spirina, E., Rivkina, E., Liang, R.,
992 Onstott, T.C., Vishnivetskaya, T., Lloyd, K.G. (2021) Eight metagenome-assembled
993 genomes provide evidence for microbial adaptation in 20,000 to 1,000,000-year-old
994 Siberian permafrost. *Appl Environ Microbiol*, Doi: 10.1128/AEM.00972-21.
- 995 [56] Sipes, K., Paul, R., Fine, A., Li, P., Liang, R., Boike, J., Onstott, T.C., Vishnivetskaya,
996 T.A., Schaeffer, S., Lloyd, K.G. (2022) Permafrost active layer microbes from Ny
997 Ålesund, Svalbard (79°N) show autotrophic and heterotrophic metabolisms with diverse
998 carbon-degrading enzymes. *Frontiers in Microbiology* 0, 3901, Doi:
999 10.3389/FMICB.2021.757812.
- 1000 [57] Sipes, K., Paul, R., Fine, A., Li, P., Liang, R., Boike, J., Onstott, T.C., Vishnivetskaya,
1001 T.A., Schaeffer, S., Lloyd, K.G. (2022) Permafrost active layer microbes from Ny
1002 Ålesund, Svalbard (79°N) show autotrophic and heterotrophic metabolisms with diverse
1003 carbon-degrading enzymes. *Front Microbiol* 0, 3901, Doi: 10.3389/FMICB.2021.757812.
- 1004 [58] Sipes, K., Paul, R., Onstott, T.C., Vishnivetskaya, T.A., Lloyd, K.G. (2022) Draft genome
1005 sequences of 10 *Pseudomonas* sp. isolates from the active layer of permafrost in Ny
1006 Ålesund, Svalbard, Norway. *Microbiology Resource Announcements* 11(6), Doi:
1007 10.1128/MRA.00201-22.
- 1008 [59] Stamps, B.W., Losey, N.A., Lawson, P.A., Stevenson, B.S. (2014) Genome sequence of
1009 *Thermoanaerobaculum aquaticum* MP-01T, the first cultivated member of *Acidobacteria*
1010 subdivision 23, isolated from a hot spring. *Genome Announcements* 2(3), 1–3, Doi:
1011 10.1128/genomeA.00570-14.
- 1012 [60] Sui, X., Frey, B., Yang, L., Liu, Y., Zhang, R., Ni, H., Li, M.H. (2022) Soil *Acidobacterial*
1013 community composition changes sensitively with wetland degradation in northeastern of
1014 China. *Frontiers in Microbiology* 13, 5176, Doi:
1015 10.3389/FMICB.2022.1052161/BIBTEX.
- 1016 [61] Suzuki, M.T., Giovannoni, S.J. (1996) Bias caused by template annealing in the
1017 amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental*
1018 *Microbiology* 62(2), 625–30, Doi: 10.1128/AEM.62.2.625-630.1996.
- 1019 [62] Tarnocai, C., Canadell, J.G., Schuur, E.A.G., Kuhry, P., Mazhitova, G., Zimov, S. (2009)
1020 Soil organic carbon pools in the northern circumpolar permafrost region. *Global*
1021 *Biogeochemical Cycles* 23(2), n/a-n/a, Doi: 10.1029/2008GB003327.
- 1022 [63] Tran, N.H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., Li, M.
1023 (2019) Deep learning enables de novo peptide sequencing from data-independent-
1024 acquisition mass spectrometry. *Nature Methods* 16(1), 63–6, Doi: 10.1038/S41592-018-
1025 0260-3.
- 1026 [64] Tripathi, B.M., Kim, H.M., Jung, J.Y., Nam, S., Hyeon Tae Ju., Kim, M., Lee, Y.K.
1027 (2019) Distinct taxonomic and functional profiles of the microbiome associated with
1028 different soil horizons of a moist tussock tundra in Alaska. *Frontiers in Microbiology*
1029 10(JUN), 1442, Doi: 10.3389/FMICB.2019.01442/BIBTEX.
- 1030 [65] Turetsky, M.R., Abbott, B.W., Jones, M.C., Anthony, K.W., Olefeldt, D., Schuur, E.A.G.,
1031 Grosse, G., Kuhry, P., Hugelius, G., Koven, C., Lawrence, D.M., Gibson, C., Sannel,

- A.B.K., McGuire, A.D. (2020) Carbon release through abrupt permafrost thaw. *Nature Geoscience* 13(2), 138–43, Doi: 10.1038/s41561-019-0526-0.
- [66] Tveit, A., Schwacke, R., Svenning, M.M., Ulrich, T. (2013) Organic carbon transformations in high-Arctic peat soils: Key functions and microorganisms. *ISME Journal* 7(2), 299–311, Doi: 10.1038/ismej.2012.99.
- [67] Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V. V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 2003 428:6978 428(6978), 37–43, Doi: 10.1038/nature02340.
- [68] Uritskiy, G. V., DiRuggiero, J., Taylor, J. (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6(1), 158, Doi: 10.1186/s40168-018-0541-1.
- [69] Wilhelm, R.C., Niederberger, T.D., Greer, C., Whyte, L.G. (2011) Microbial diversity of active layer and permafrost in an acidic wetland from the Canadian High Arctic. *Canadian Journal of Microbiology* 57(4), 303–15, Doi: 10.1139/w11-004.
- [70] Wilson, A., Norden, N. (2015) The R Project for Statistical Computing The R Project for Statistical Computing. URL: [Http://Www. r-Project. Org/254](http://www.r-project.org/254). <https://www.r-project.org/>. [accessed May 3, 2023].
- [71] Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A.F.F., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B., Wilson, R.M., Purvine, S.O., Nicora, C.D., Li, C., Froelking, S., Chanton, J.P., Crill, P.M., Saleska, S.R., Rich, V.I., Tyson, G.W. (2018) Genome-centric view of carbon processing in thawing permafrost. *Nature* 560(7716), 49–54, Doi: 10.1038/s41586-018-0338-1.
- [72] Wu, X., Almatari, A.L., Cyr, W.A., Williams, D.E., Pfiffner, S.M., Rivkina, E.M., Lloyd, K.G., Vishnivetskaya, T.A. (2023) Microbial life in 25-m-deep boreholes in ancient permafrost illuminated by metagenomics. *Environmental Microbiome* 2023 18:1 18(1), 1–19, Doi: 10.1186/S40793-023-00487-9.
- [73] Wu, X., Chauhan, A., Layton, A.C., Lau Vetter, M.C.Y., Stackhouse, B.T., Williams, D.E., Whyte, L., Pfiffner, S.M., Onstott, T.C., Vishnivetskaya, T.A. (2021) Comparative Metagenomics of the Active Layer and Permafrost from Low-Carbon Soil in the Canadian High Arctic. *Environmental Science and Technology* 55(18), 12683–93, Doi: 10.1021/ACS.EST.1C00802/SUPPL_FILE/ES1C00802_SI_001.PDF.
- [74] Wu, Y.-W., Simmons, B.A., Singer, S.W. (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32(4), 605–7, Doi: 10.1093/bioinformatics/btv638.
- [75] Xue, Y., Jonassen, I., Øvreås, L., Taş, N. (2020) Metagenome-assembled genome distribution and key functionality highlight importance of aerobic metabolism in Svalbard permafrost. *FEMS Microbiology Ecology* 96(5), Doi: 10.1093/FEMSEC/FIAA057.
- [76] Yergeau, E., Bokhorst, S., Kang, S., Zhou, J., Greer, C.W., Aerts, R., Kowalchuk, G.A. (2012) Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *The ISME Journal* 6(3), 692–702, Doi: 10.1038/ismej.2011.124.
- [77] Yi, Y., Kimball, J.S., Chen, R.H., Moghaddam, M., Reichle, R.H., Mishra, U., Zona, D., Oechel, W.C. (2018) Characterizing permafrost active layer dynamics and sensitivity to landscape spatial heterogeneity in Alaska. *Cryosphere* 12(1), 145–61, Doi: 10.5194/TC-12-145-2018.

[78] Zhang, X., Dong, W., Dai, X., Schaeffer, S., Yang, F., Radosevich, M., Xu, L., Liu, X., Sun, X. (2015) Responses of absolute and specific soil enzyme activities to long term additions of organic and mineral fertilizer. *Science of the Total Environment* 536, 59–67, Doi: 10.1016/j.scitotenv.2015.07.043.

Figures for NovaSeq Sipes Paper:

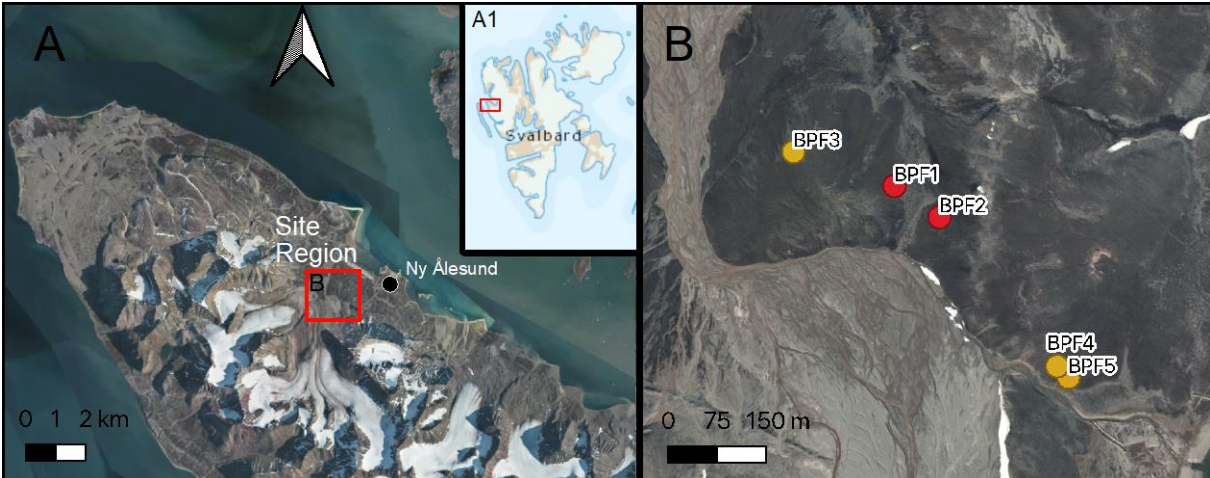


Figure 1: Sampling location near the research station, Ny Ålesund (A) on the Svalbard archipelago (A1), close to the Bayelva long term permafrost observatory (B – fence can be seen above ‘BPF3 marker’). Site region is expanded in panel B. Samples collected in April 2018 are in red and are investigated in Sipes et al, 2022, samples collected in September 2019 are in gold (this study). Satellite photos are from Norwegian Polar Institute Map Data and Services and map was rendered in QGIS v3.18.

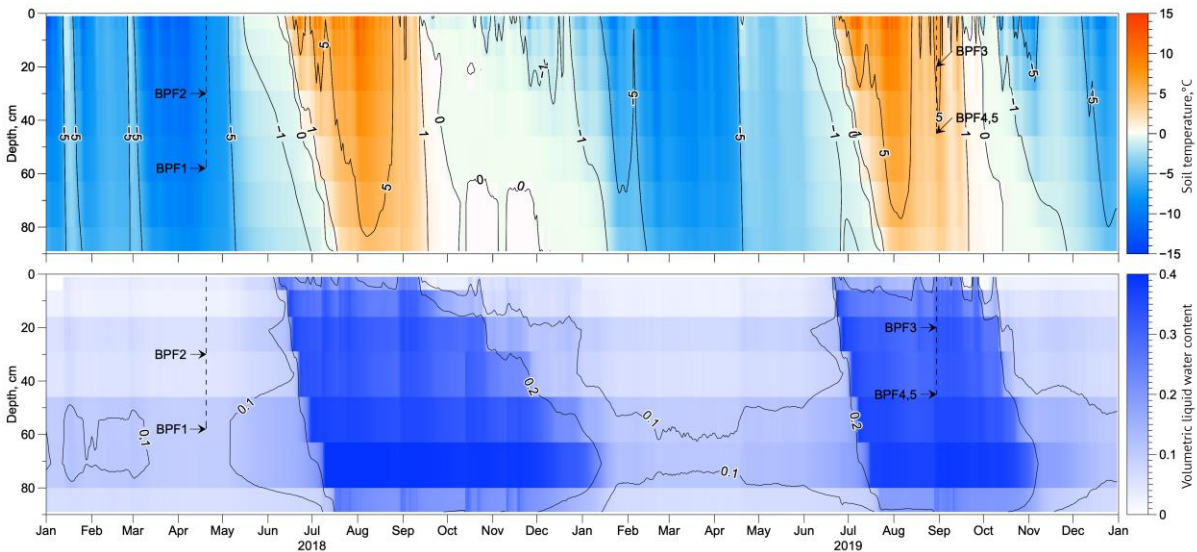


Figure 2: Contour plots displaying the (top) soil temperature and (bottom) volumetric liquid water content profiles for the soil at the Bayelva site between January 2018 to December 2019, mean daily values are based on hourly data collected at depths of 1, 11, 21, 37, 55, 71, and 89 cm below the surface. The dates of sampling of soil material in frozen (April 2018) and unfrozen (September 2019) are added to the graph.

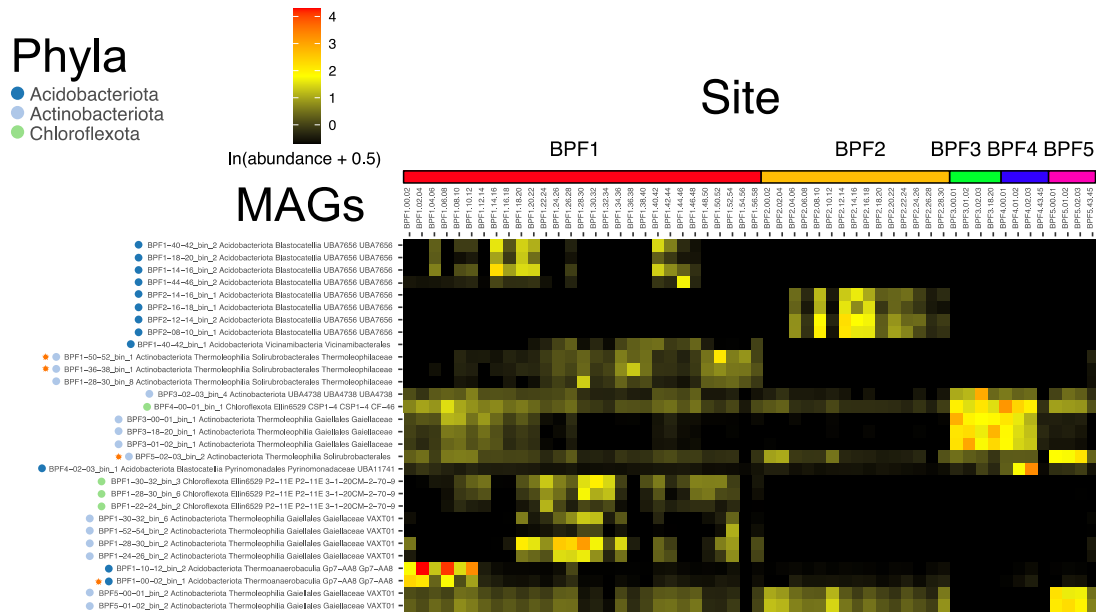


Figure 3: Heatmap for read abundance (genome copies per million length-weighted average of the bin's contig abundances) for 30 representatives of the total MAGs (n=169) against the assembled metagenomes (n=56). Shown are the 30 MAGs from the full heatmap (Figure S5) that represent the most abundant MAGs from each of the three major clusters on the hierarchical clustering shown in Figure S5. Abundance plus 0.5 has been natural log transformed for visual representation on a positive scale. Sites are ordered left to right from surface to bottom depth. Colored dot at beginning of the MAG name depicts phylum. Orange stars next to the MAG name denote the present of autotrophic metabolic pathways. All autotrophic MAG abundance shown in Figure S2. Heatmap depiction inspired from others [51].

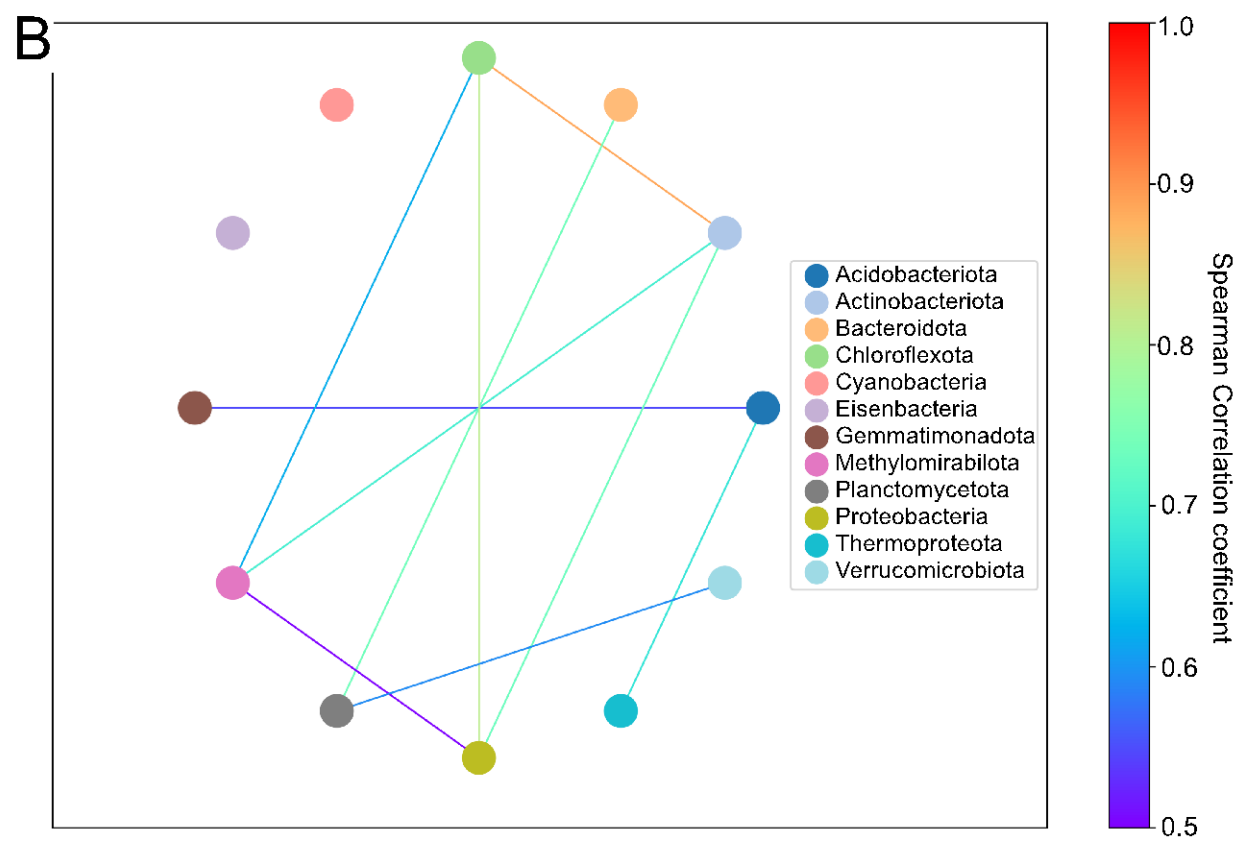
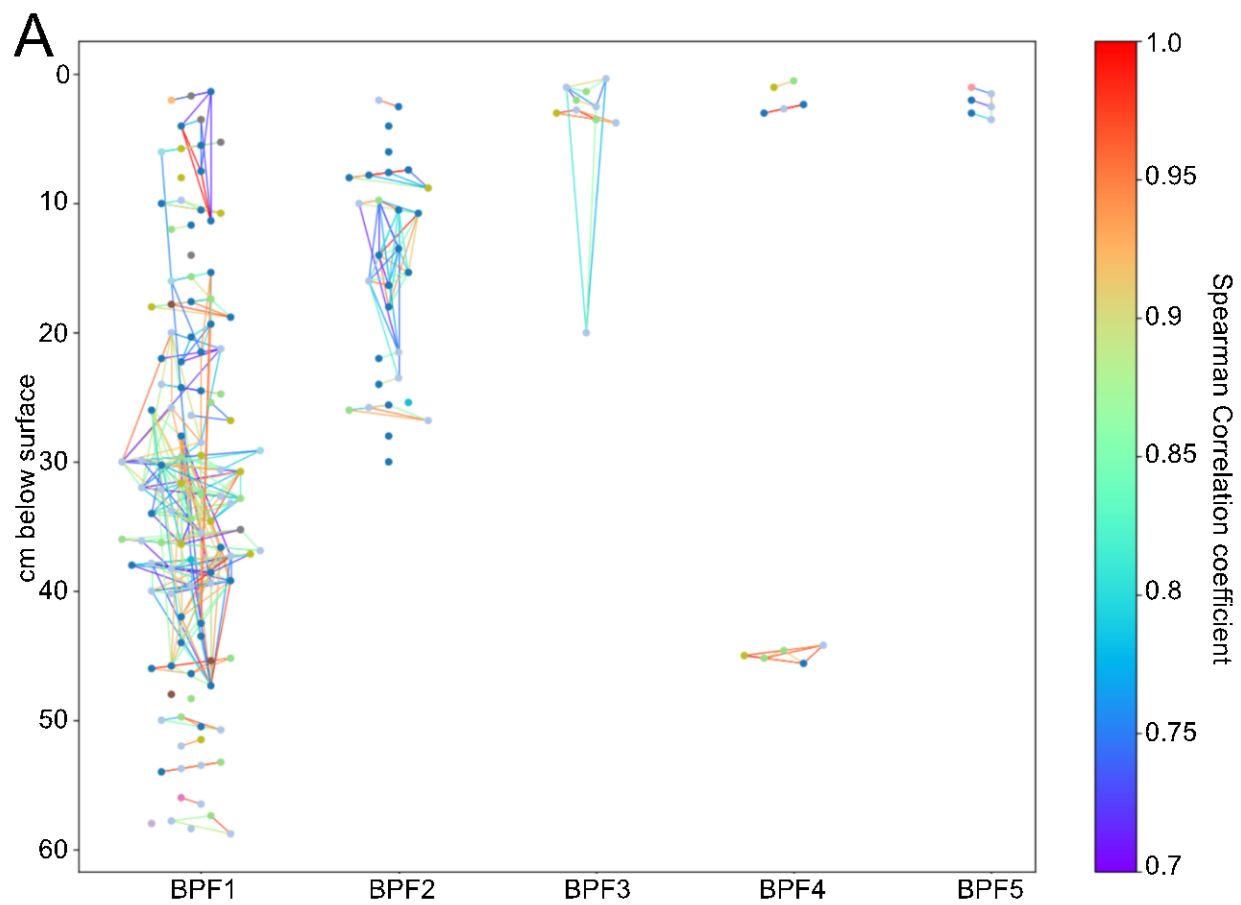


Figure 4: Network analysis A) Each MAG correlation based on the unique read abundance across all samples. Spearman correlation coefficient is represented with the line between MAGs and corresponds to the colors in the legend. MAGs originating from the same location are jittered along the 2 cm origin source in order to visualize correlations between MAGs from the same location B) Phylum level co-correlation relationships based on read abundances across all samples.

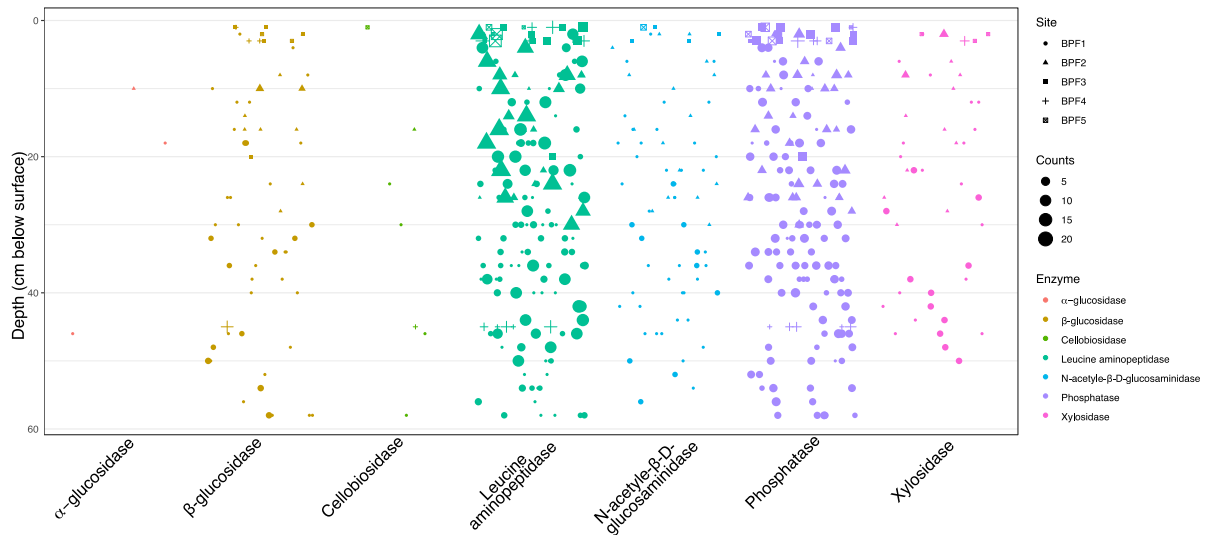


Figure 5. Gene counts for carbohydrate- and peptide-degrading enzyme per MAG by depth and site. Number of single genes that encode for an enzyme is represented by the size of the symbol, color represents the enzyme type, and shape represents the sample site. Number of individual enzyme-coding genes present in each MAG can be found in Table S4.

Table 1. Summary of freeze back characteristics (temperature-T, volumetric water content- VWC) analyzed from the Bayelva site. The start of freeze back was defined at drop of temperatures below -0.1°C and continuously remaining below -0.1°C. The duration of the period of zero curtain was defined as period constantly around 0°C until the hourly soil temperature dropped and remained below -0.5°C. The mean remaining VWC was calculated for 24 hours for the last day of the period of zero curtain.

Dep th of sens or (cm)	Drop below 0.1°C (start freeze back)	Duration of freeze back (zero curtain, soil temp. <-0.5°C)	Dura tion of freez e back (days)	Mean Remaini ng VWC at end of freeze back (%)	Mean soil temp. April 15. 2018 (°C)	Mean VWC April 15. 2018 (%)	Mean soil temp. Septembe r 15. 2019 (°C)	Mean VWC Septembe r 15. 2019 (%)
1	4.10.2017	4. - 24.10.2017	20	2.4	-9	4.3	3.6	20
11	7.10.2017	7. – 25.10.2017	18	10.8	-8.9	4.5	3.8	29.1

21	8.10.2017	8. – 30.10.2017	22	11.2	-8.9	6.9	3.7	32.4
37	11.10.2017	11.10. – 3.11.2017	23	12.4	-9	5.9	3.4	32.4
55	14.10.2017	14.10. – 9.11.2017	26	18.7	-8.8	9.5	3.1	35.4
71	22.10.2017	22.10. - 21.11.2017	30	22.6	-8.6	10.5	2.6	38.9
89	22.10.2017	22.10. – 26.11.2017	35	10.5	-8.1	6.6	2	21.2

1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178

Supplemental Tables and Figures for: Depth-specific distribution of bacterial MAGs in permafrost active layer in Ny Ålesund, Svalbard (79°N)

Katie Sipes^{a*#}, Joy Buongiorno^a, Andrew D. Steen^{a,b}, Andrey A. Abramov^c, Samantha L. Peters^d, Richard J. Gianonne^d, Robert L. Hettich^d, Julia Boike^{e,f}, Sarahi L. Garcia^h, Tatiana A. Vishnivetskaya^a, Karen G. Lloyd^a

^aDepartment of Microbiology, University of Tennessee, Knoxville, United States

^bDepartment of Earth and Planetary Sciences, University of Tennessee, Knoxville, United States

^cSoil Cryology Laboratory, Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences, Pushchino, Russia

^dBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

^eAlfred Wegener Institute for Polar and Marine Research, Potsdam, Germany

^fDepartment of Geography, Humboldt University, Berlin, Germany

^gDepartment of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

^h*Current affiliation: Department of Environmental Sciences, Aarhus University, 4000 Roskilde, Denmark

Corresponding author: ksipes@envs.au.dk

Running title: Depth distribution of MAGs in Ny Ålesund, Svalbard

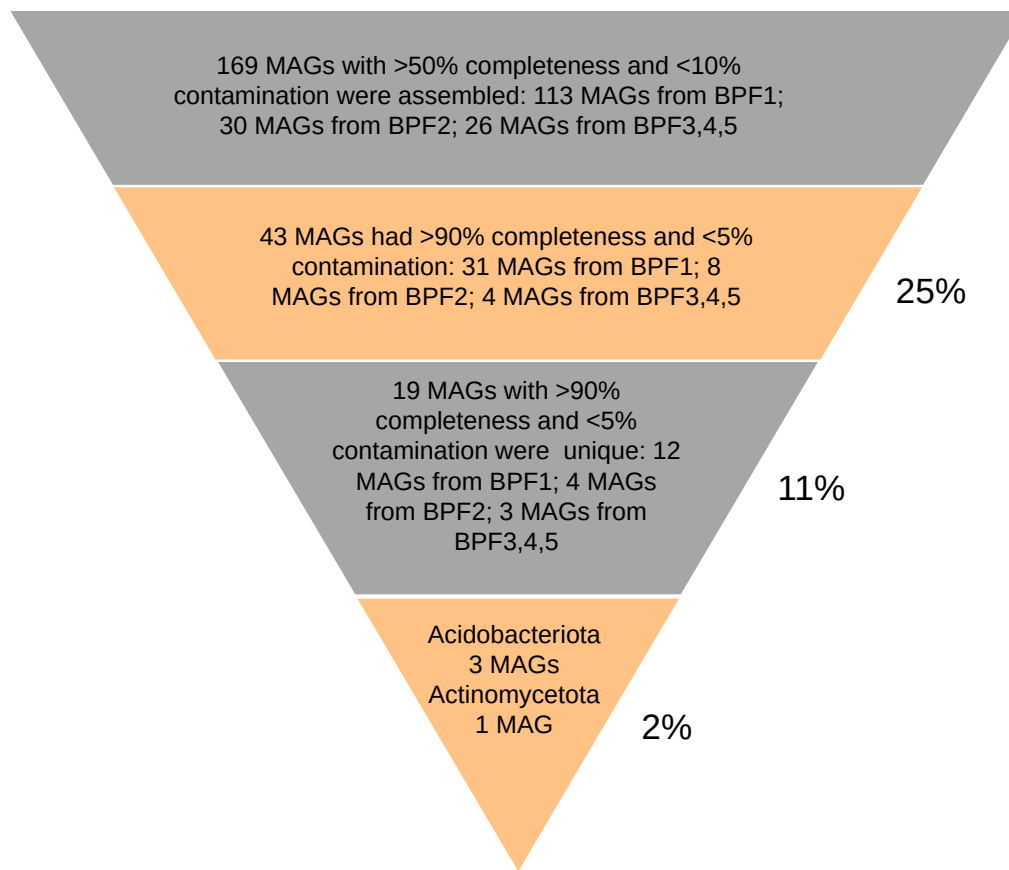


Fig. S1. The diagram shows how application of stringency conditions decreases number of MAGs while increasing MAG quality. Out of 169 MAGs, only 3 Acidobacteriota and 1 Actinomycetota MAGs had >75% 16S rRNA gene completeness and number of contigs less than 100 with >80% of tRNAs present. Two Acidobacteriota MAGs, which are in the same order, have <78.6% ANI between them.

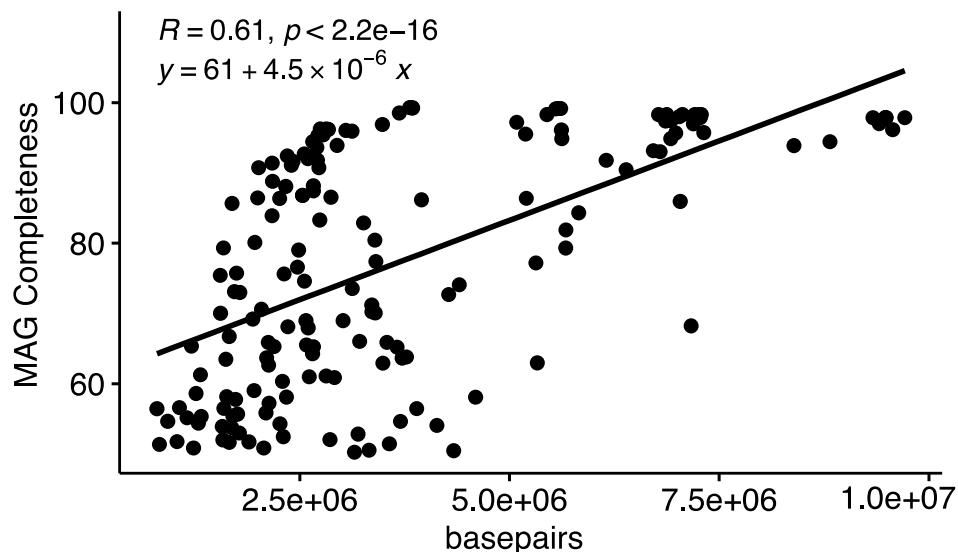


Figure S2: MAG completeness and total base pairs with regression line and R² value.

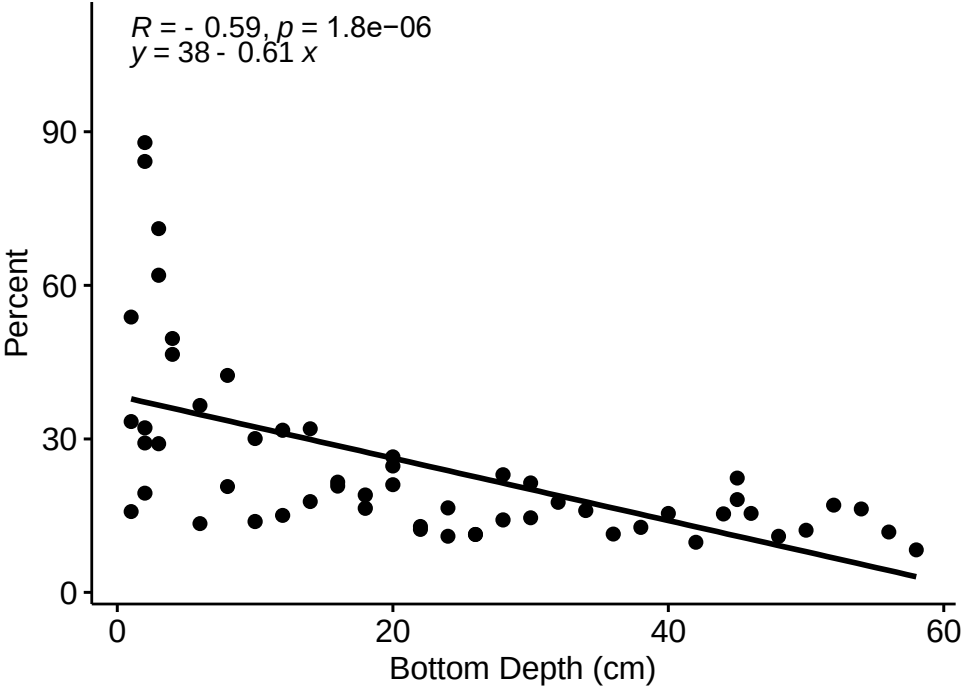


Figure S3: Cumulative percent of an assembly's reads that mapped to any of the 169 MAGs in this study compared to the bottom depth of the sample assembly with regression line and R² value.

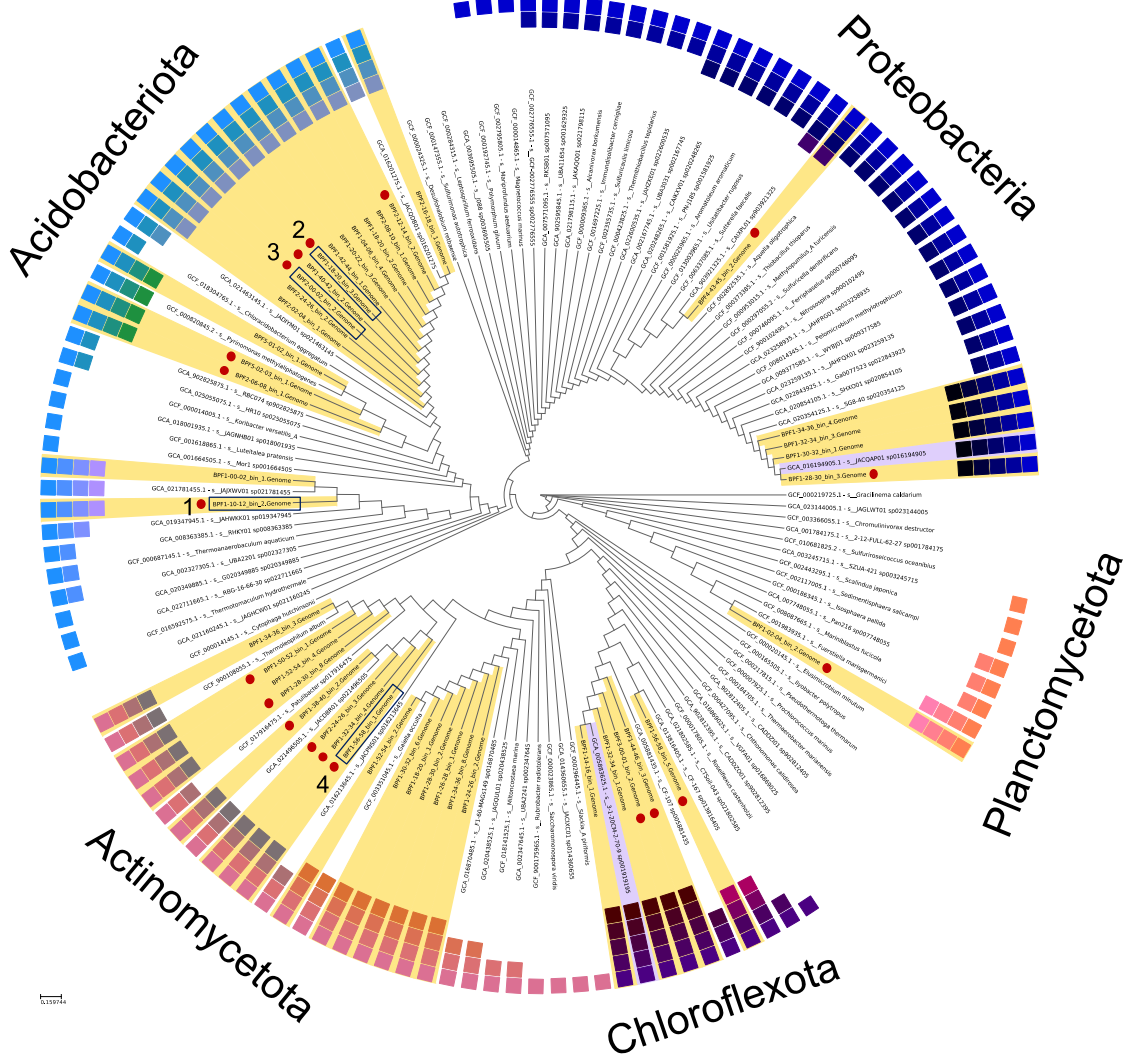


Fig. S4. The genome tree of the 43 MAGs with >90% completeness and <5% contamination (highlighted in yellow) classified using GTDB TK v.1.7.0 program. MAGs marked with red circles represent 19 unique MAGs with <95% ANI and were uploaded to the NCBI under BioProject ID number PRJNA1040778. MAGs under numbers 1, 2, 3, and 4 have >75% 16S rRNA gene completeness, and number of contigs less than 100 with >80% of tRNAs present. These MAGs were selected for the SeqCode submission.

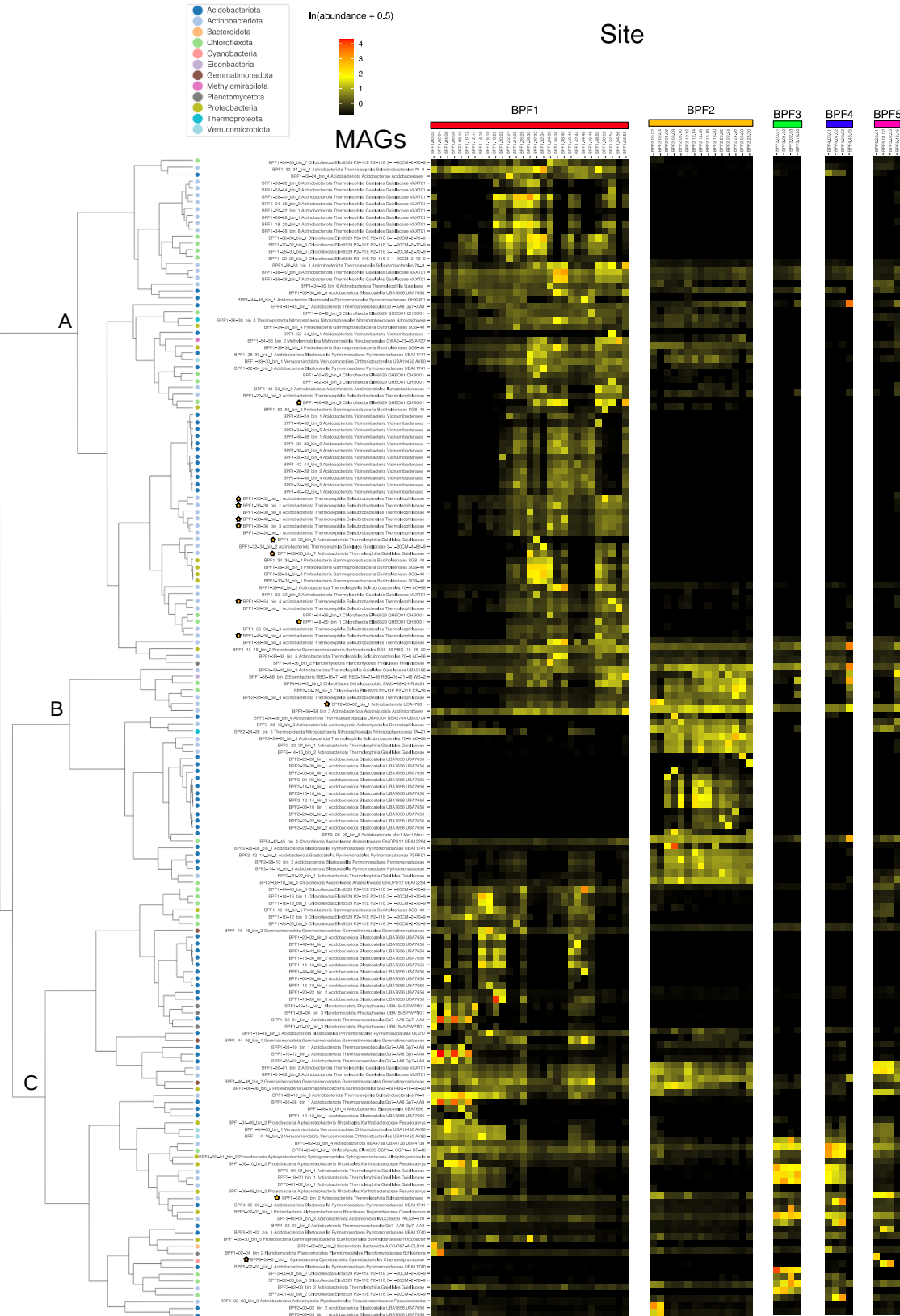
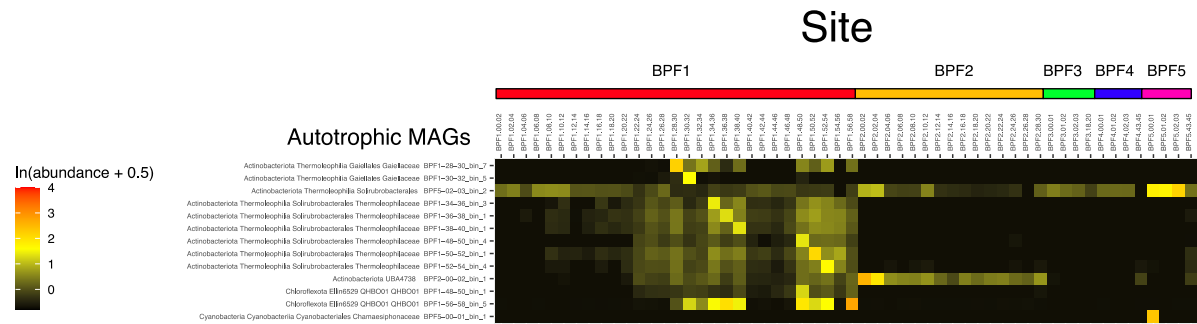


Figure S5: Heatmap for reads per million mapped reads for MAGs (n=169) to the assembled metagenomes (n=56). Abundance plus 0.5 has been natural log transformed for visual representation on a positive scale. MAGs are clustered based on spearman correlation coefficient and show family level taxonomy when possible. Letters at the branches of the dendrogram signify observed clustering patterns: A) A population predominantly at the bottom half of BPF1, B) a predominantly BPF2 population and C) a surface BPF1 population and at the three summer thawed sites. Sites are ordered from surface to bottom depth. Colored dot at the end of the correlation dendrogram shows phylum. Brown stars next to the MAG name denote the present of autotrophic metabolic pathways. Heatmap depiction inspired from others (Rogers et al., 2022).



A



B

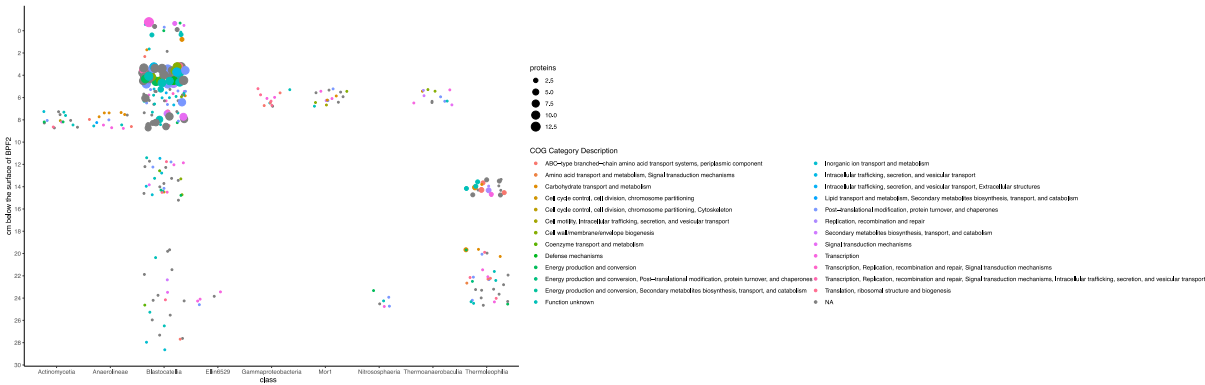


Figure S7: Proteomics Downcore of BPF2. Divided by identified A) phyla or B) class and the protein’s COG category. “NA” and “None” can be considered the same, while “Function Unknown” is truly unknown. The metaproteome was produced from the 15-30 cm section of BPF 2, and the figure shows the abundance of MAGs from all the depths in BPF2 that map to these proteins.

Table S1: Metagenomic information for each layer after NovaSeq sequencing

Location	Top depth	Bottom depth	Unassembled basepair	Assembled basepairs	Total contigs
BPF1	0	2	123752000	21379400	220684
BPF1	2	4	511550000	25196864	256501
BPF1	4	6	257789000	44972143	465859
BPF1	6	8	494146000	23733448	245156
BPF1	8	10	327207000	37219008	385244

BPF1	10	12	394310000	37767829	392037
BPF1	12	14	220500000	32194756	335798
BPF1	14	16	393295000	48117185	498103
BPF1	16	18	487565000	80263829	831809
BPF1	18	20	663201000	55131041	569093
BPF1	20	22	602192000	79345546	816243
BPF1	22	24	637781000	81064796	837128
BPF1	24	26	836668000	79791659	823051
BPF1	26	28	427391000	41121782	423285
BPF1	28	30	743619000	84094194	864994
BPF1	30	32	717598000	73408580	754482
BPF1	32	34	746948000	82083027	844318
BPF1	34	36	1050380000	108045790	1109676
BPF1	36	38	945814000	107063231	1101565
BPF1	38	40	917852000	90011304	926645
BPF1	40	42	748074000	90693688	935989
BPF1	42	44	434197000	56352451	580333
BPF1	44	46	591580000	77247068	797759
BPF1	46	48	651528000	74954810	775080
BPF1	48	50	857033000	86912274	893124
BPF1	50	52	503281000	42416139	438922
BPF1	52	54	598423000	50325735	518904
BPF1	54	56	549275000	34468968	358714
BPF1	56	58	1202870000	132323551	1363834
BPF2	0	2	293181000	35396041	364457
BPF2	2	4	131723000	20287310	209841
BPF2	4	6	495011000	56848658	587131
BPF2	6	8	592161000	62035095	638901
BPF2	8	10	605284000	71566089	739609
BPF2	10	12	323945000	44852417	466677
BPF2	12	14	480953000	47926171	496720
BPF2	14	16	396312000	40457661	417934
BPF2	16	18	365513000	41227452	426740
BPF2	18	20	215474000	28053124	291270
BPF2	20	22	549503000	62297018	643745
BPF2	22	24	419000000	37687223	388878
BPF2	24	26	699134000	82221231	850568
BPF2	26	28	388181000	27573769	286386
BPF2	28	30	397360000	46230632	477580

BPF3	0	1	410058000	43238933	448394
BPF3	1	2	281359000	27651377	286728
BPF3	2	3	183225000	21293223	220426
BPF3	18	20	320237000	33402913	347236
BPF4	0	1	164832000	20197735	211571
BPF4	1	2	74010500	10797152	113843
BPF4	2	3	171944000	26173812	269925
BPF4	43	45	787358000	48415591	503305
BPF5	0	1	357556000	43519401	450761
BPF5	1	2	306314000	46678320	482117
BPF5	2	3	230418000	34768486	359435
BPF5	43	45	173888000	20710178	216942

1276
1277
1278

1279 Table S2: Information on all 169 MAGs' completeness and contamination percentages, phylum,
1280 number of reads and total number of base pairs. The MAG column displays the site and the depth
1281 increment the MAG was assembled from.

MAG	Completeness	Contamination	Phyla	Number of reads	Base pairs
BPF1-00-02_bin_1	97.2	4.27	Acidobacteriota	83724	5085009
BPF1-00-02_bin_2	72.69	4.54	Bacteroidota	62024	4275032
BPF1-00-02_bin_3	54.64	1.14	Planctomycetota	53907	3699313
BPF1-02-04_bin_1	99.15	7.12	Acidobacteriota	91368	5569227
BPF1-02-04_bin_2	93.88	3.37	Planctomycetota	137848	8390584
BPF1-04-06_bin_1	55.68	2.41	Verrucomicrobiota	29403	1757510
BPF1-04-06_bin_2	60.88	5.66	Proteobacteria	48665	2912269
BPF1-04-06_bin_3	58.09	5.88	Planctomycetota	76502	4594825
BPF1-04-06_bin_4	97.39	3.89	Acidobacteriota	113938	6924428
BPF1-06-08_bin_1	99.15	5.41	Acidobacteriota	91971	5608833

BPF1-06-08_bin_2	61.1	1.27	Proteobacteria	47017	2813973
BPF1-08-10_bin_1	63.65	3.92	Acidobacteriota	61596	3718528
BPF1-08-10_bin_2	86.16	3.9	Proteobacteria	65448	3950780
BPF1-08-10_bin_3	70.59	7.47	Actinobacteriota	33679	2041770
BPF1-08-10_bin_4	54.06	4.6	Acidobacteriota	60402	4133684
BPF1-10-12_bin_1	56.45	3.53	Acidobacteriota	65075	3892517
BPF1-10-12_bin_2	99.1	4.56	Acidobacteriota	91023	5550592
BPF1-10-12_bin_3	65.28	7.27	Chloroflexota	31617	2192232
BPF1-12-14_bin_1	74.08	5.75	Planctomycetota	73213	4401270
BPF1-14-16_bin_1	91.09	0.99	Chloroflexota	39442	2398630
BPF1-14-16_bin_2	98.29	5.56	Acidobacteriota	119660	7287078
BPF1-14-16_bin_3	62.63	8.78	Verrucomicrobiota	35574	2127242
BPF1-16-18_bin_1	86.36	0	Chloroflexota	37177	2258351
BPF1-16-18_bin_2	80.42	9.34	Gemmatimonadota	56042	3395095
BPF1-16-18_bin_3	51.65	0	Acidobacteriota	27731	1658211
BPF1-16-18_bin_4	93.02	5.77	Acidobacteriota	111747	6795176
BPF1-16-18_bin_5	51.44	5.77	Proteobacteria	52163	3567515
BPF1-18-20_bin_1	95.32	0.86	Actinobacteriota	44569	2715467
BPF1-18-20_bin_2	98.29	3.42	Acidobacteriota	115954	7061658
BPF1-18-20_bin_3	98.01	3.42	Acidobacteriota	115222	7024570
BPF1-20-22_bin_1	87.43	1.72	Actinobacteriota	43764	2662467

BPF1-20-22_bin_2	70.23	4.27	Acidobacteriota	55648	3358736
BPF1-20-22_bin_3	93.16	2.99	Acidobacteriota	110235	6714501
BPF1-20-22_bin_4	68.23	3.42	Acidobacteriota	117715	7164573
BPF1-22-24_bin_1	62.97	2.56	Acidobacteriota	87587	5330317
BPF1-22-24_bin_2	74.59	1.98	Chloroflexota	42048	2554810
BPF1-22-24_bin_3	88.07	2.13	Actinobacteriota	33321	2331863
BPF1-22-24_bin_4	57.76	1.07	Acidobacteriota	24938	1733386
BPF1-24-26_bin_1	75.73	0.86	Actinobacteriota	28778	1744943
BPF1-24-26_bin_2	96.26	1.72	Actinobacteriota	39697	2812353
BPF1-24-26_bin_3	66.72	4.05	Chloroflexota	27423	1657415
BPF1-24-26_bin_4	55.42	7.12	Proteobacteria	28475	1702222
BPF1-24-26_bin_5	97.86	8.6	Acidobacteriota	155665	9485262
BPF1-26-28_bin_1	95.4	0.86	Actinobacteriota	45588	2778077
BPF1-26-28_bin_2	97.77	8.6	Acidobacteriota	155612	9479659
BPF1-28-30_bin_1	58.61	3.21	Verrucomicrobiota	21135	1262838
BPF1-28-30_bin_2	96.26	1.29	Actinobacteriota	45001	2743391
BPF1-28-30_bin_3	99.26	2.84	Proteobacteria	63155	3845301
BPF1-28-30_bin_4	58.1	0.95	Acidobacteriota	39020	2339490
BPF1-28-30_bin_5	63.8	1.09	Proteobacteria	62891	3770135
BPF1-28-30_bin_6	83.29	0	Chloroflexota	44936	2737534
BPF1-28-30_bin_7	68.97	1.72	Actinobacteriota	42267	2572961

BPF1-28-30_bin_8	91.39	2.76	Actinobacteriota	35799	2169401
BPF1-30-32_bin_1	99.29	2.37	Proteobacteria	62606	3810752
BPF1-30-32_bin_2	54.4	7.76	Actinobacteriota	21369	1289837
BPF1-30-32_bin_3	88.17	0	Chloroflexota	43685	2661212
BPF1-30-32_bin_4	56.6	3.44	Chloroflexota	17774	1063606
BPF1-30-32_bin_5	59.03	8.5	Actinobacteriota	32314	1954520
BPF1-30-32_bin_6	96.18	1.55	Actinobacteriota	40099	2840431
BPF1-32-34_bin_1	95.93	1.49	Chloroflexota	44065	3122888
BPF1-32-34_bin_2	51.72	3.45	Actinobacteriota	31167	1893215
BPF1-32-34_bin_3	99.22	2.37	Proteobacteria	63081	3840918
BPF1-32-34_bin_4	95.52	2.44	Actinobacteriota	38874	2749473
BPF1-32-34_bin_5	65.88	8.46	Acidobacteriota	51299	3538091
BPF1-34-36_bin_1	73.1	1.85	Chloroflexota	28315	1720263
BPF1-34-36_bin_2	85.94	4.79	Planctomycetota	116171	7035207
BPF1-34-36_bin_3	92.7	0.86	Actinobacteriota	42060	2551783
BPF1-34-36_bin_4	98.52	2.84	Proteobacteria	60557	3685288
BPF1-34-36_bin_5	97.86	8.6	Acidobacteriota	155600	9482575
BPF1-34-36_bin_6	65.34	1.72	Actinobacteriota	19887	1207756
BPF1-34-36_bin_7	61.27	2.78	Chloroflexota	21738	1314141
BPF1-34-36_bin_8	92.02	0.86	Actinobacteriota	42613	2595568
BPF1-36-38_bin_1	86.44	1.81	Actinobacteriota	32919	1997020

BPF1-36-38_bin_2	97.86	8.6	Acidobacteriota	153134	9332488
BPF1-36-38_bin_3	86.79	5.87	Proteobacteria	41914	2532115
BPF1-36-38_bin_4	56.44	0	Thermoproteota	13233	796143
BPF1-36-38_bin_5	55.16	0.6	Actinobacteriota	19015	1154487
BPF1-36-38_bin_6	50.54	5.18	Acidobacteriota	55505	3325559
BPF1-36-38_bin_7	65.26	2.59	Actinobacteriota	44110	2662034
BPF1-38-40_bin_1	88.79	0.86	Actinobacteriota	35865	2174848
BPF1-38-40_bin_2	93.91	1.75	Actinobacteriota	48404	2943805
BPF1-38-40_bin_3	86.54	1.72	Actinobacteriota	47352	2870636
BPF1-38-40_bin_4	60.34	0	Actinobacteriota	37757	2291851
BPF1-38-40_bin_5	96.15	8.6	Acidobacteriota	156980	9568184
BPF1-40-42_bin_1	97.86	9.03	Acidobacteriota	155606	9480392
BPF1-40-42_bin_2	94.87	3.85	Acidobacteriota	113623	6922424
BPF1-42-44_bin_1	97.38	3.85	Acidobacteriota	112796	6859524
BPF1-42-44_bin_2	97.01	8.72	Acidobacteriota	154441	9407608
BPF1-44-46_bin_1	71.19	3.4	Gemmatimonadota	55594	3357764
BPF1-44-46_bin_2	94.87	6.81	Acidobacteriota	92870	5622980
BPF1-44-46_bin_3	90.74	0	Chloroflexota	44741	2727255
BPF1-44-46_bin_4	97.86	8.93	Acidobacteriota	137048	9712036
BPF1-44-46_bin_5	52.46	4.32	Acidobacteriota	38466	2302678
BPF1-46-48_bin_1	97.86	8.6	Acidobacteriota	155714	9488971

BPF1-46-48_bin_2	64.26	2.59	Gemmatimonadota	44123	2651292
BPF1-46-48_bin_3	63.48	8.9	Chloroflexota	23400	1618096
BPF1-48-50_bin_1	85.65	0.31	Chloroflexota	27828	1691584
BPF1-48-50_bin_2	51.77	0.85	Actinobacteriota	17441	1035366
BPF1-48-50_bin_3	94.44	8.6	Acidobacteriota	144995	8820072
BPF1-48-50_bin_4	86.81	2.01	Actinobacteriota	36030	2532218
BPF1-50-52_bin_1	91.65	1.38	Actinobacteriota	39886	2419480
BPF1-50-52_bin_2	52.98	3.95	Proteobacteria	29764	1778409
BPF1-52-54_bin_1	50.86	0	Acidobacteriota	34218	2069668
BPF1-52-54_bin_2	94.44	3.82	Actinobacteriota	43655	2652387
BPF1-52-54_bin_3	79.32	2.44	Chloroflexota	26260	1588263
BPF1-52-54_bin_4	91.81	0.86	Actinobacteriota	38534	2709567
BPF1-54-56_bin_1	83.91	2.16	Actinobacteriota	35743	2168358
BPF1-54-56_bin_2	55.85	9.49	Methylomirabilota	34893	2096020
BPF1-56-58_bin_1	92.41	1.29	Actinobacteriota	38608	2350257
BPF1-56-58_bin_2	65.52	0	Eisenbacteria	42398	2578006
BPF1-56-58_bin_3	79.02	2.14	Actinobacteriota	41099	2485213
BPF1-56-58_bin_4	75.42	3.02	Actinobacteriota	25573	1551321
BPF1-56-58_bin_5	90.74	0	Chloroflexota	28342	2009185
BPF2-00-02_bin_1	75.61	0.85	Actinobacteriota	38374	2312313
BPF2-00-02_bin_2	98.29	3.42	Acidobacteriota	112749	6873804

BPF2-02-04_bin_1	90.44	3.85	Acidobacteriota	105408	6390013
BPF2-04-06_bin_1	98.29	6.31	Acidobacteriota	111363	6777189
BPF2-06-08_bin_1	95.67	3.42	Acidobacteriota	98521	6980728
BPF2-06-08_bin_2	63.68	1.4	Proteobacteria	35207	2103368
BPF2-06-08_bin_3	79.31	8.93	Acidobacteriota	93846	5669428
BPF2-06-08_bin_4	52.83	5.09	Acidobacteriota	46272	3194568
BPF2-06-08_bin_5	66.03	7.76	Acidobacteriota	45951	3212389
BPF2-08-10_bin_1	98.29	3.42	Acidobacteriota	118313	7209516
BPF2-08-10_bin_2	73.53	0.85	Acidobacteriota	51913	3124438
BPF2-08-10_bin_3	70.06	9.48	Actinobacteriota	49387	3398773
BPF2-08-10_bin_4	50.25	5.74	Chloroflexota	45929	3151136
BPF2-12-14_bin_1	60.98	2.56	Acidobacteriota	43562	2611467
BPF2-12-14_bin_2	98.29	4.27	Acidobacteriota	119150	7261543
BPF2-14-16_bin_1	95.73	6.61	Acidobacteriota	120135	7316180
BPF2-14-16_bin_2	68.1	0	Actinobacteriota	38803	2356270
BPF2-14-16_bin_3	52.04	2.59	Acidobacteriota	47635	2859119
BPF2-16-18_bin_1	97.86	4.27	Acidobacteriota	119378	7273358
BPF2-20-22_bin_1	52	1.87	Actinobacteriota	26160	1582933
BPF2-20-22_bin_2	96.98	7.23	Acidobacteriota	118222	7191273
BPF2-22-24_bin_1	76.58	1.2	Actinobacteriota	40755	2471623
BPF2-22-24_bin_2	98.2	5.24	Acidobacteriota	118519	7205907

BPF2-24-26_bin_1	53.91	0.93	Chloroflexota	25969	1575882
BPF2-24-26_bin_2	91.79	4.61	Acidobacteriota	101330	6153191
BPF2-24-26_bin_3	93.62	2.73	Actinobacteriota	44415	2702881
BPF2-24-26_bin_4	55.34	0	Actinobacteriota	21952	1323665
BPF2-24-26_bin_5	54.65	3.88	Thermoproteota	13550	925091
BPF2-26-28_bin_1	81.88	8.08	Acidobacteriota	94178	5672543
BPF2-28-30_bin_1	84.32	5.45	Acidobacteriota	96334	5825415
BPF3-00-01_bin_1	80.1	0.57	Actinobacteriota	32416	1963645
BPF3-00-01_bin_2	96.04	0	Chloroflexota	49981	3048075
BPF3-00-01_bin_3	68.97	0	Actinobacteriota	49749	3015741
BPF3-01-02_bin_1	72.99	0	Actinobacteriota	29378	1783989
BPF3-01-02_bin_2	65.89	0	Chloroflexota	34890	2122836
BPF3-02-03_bin_1	65.2	4.99	Proteobacteria	61151	3659402
BPF3-02-03_bin_2	69.2	6.58	Actinobacteriota	32027	1938534
BPF3-02-03_bin_3	56.49	0	Chloroflexota	26245	1593500
BPF3-02-03_bin_4	53.73	0.85	Actinobacteriota	28264	1688417
BPF3-18-20_bin_1	70.03	0.93	Actinobacteriota	25600	1554212
BPF4-00-01_bin_1	67.95	7.41	Chloroflexota	37376	2599343
BPF4-00-01_bin_2	54.29	7.07	Proteobacteria	32928	2262842
BPF4-02-03_bin_1	96.1	5.3	Acidobacteriota	92493	5618098
BPF4-02-03_bin_2	62.92	1.52	Acidobacteriota	58308	3490178

BPF4-02-03_bin_3	77.18	3.67	Actinobacteriota	88237	5314118
BPF4-43-45_bin_1	77.39	0.85	Acidobacteriota	56565	3406324
BPF4-43-45_bin_2	96.89	2.92	Proteobacteria	57351	3483897
BPF4-43-45_bin_3	50.86	0	Actinobacteriota	20314	1232642
BPF4-43-45_bin_4	82.88	2	Chloroflexota	54220	3259648
BPF4-43-45_bin_5	51.36	1.6	Chloroflexota	13815	825344
BPF5-00-01_bin_1	86.37	4.52	Cyanobacteria	75200	5199531
BPF5-00-01_bin_2	57.23	5.8	Actinobacteriota	30749	2132456
BPF5-01-02_bin_1	95.51	2.61	Acidobacteriota	85618	5191136
BPF5-01-02_bin_2	58.16	6.77	Actinobacteriota	23313	1626993
BPF5-02-03_bin_1	98.29	2.56	Acidobacteriota	89424	5445790
BPF5-02-03_bin_2	50.47	9.48	Actinobacteriota	63038	4334012

Table S3: Linear regression between the percent of assembled reads binned per sample and the number of MAGs, MAG size or MAG completeness.

	Percentage of assembled reads binned per sample
Number of MAGs per sample	$R^2 = -0.031$ $p = 0.69$
MAG completeness	$R^2 = 0.012$ $p = 0.87$
MAG size (bp)	$R^2 = 0.2$ $p = 0.0082$

Table S4: Percentage of reads in metagenomic assemblies identified in each phylum by kaiju
 parsed out by boreholes. Only a representative subset of samples was included in this analysis, as
 it is a check on the phyla identified by MAGs. The representative samples used for this analysis
 were BPF1-23-30, BPF1-50-52, BPF3-1-2, BPF3-18-20, and BPF5-43-45.

Phylum	BPF1	BPF3	BPF5
Acidobacteria	2.86	1.33	2.74
Actinobacteria		60.19	52.05
Ascomycota		2.01	3.11
Bacteroidetes		1.74	3.53
Chloroflexota	26.14	4.04	5.43
Cyanobacteria		1.35	3.24
FCB	8.11		
Firmicutes		3.82	3.69
Proteobacteria	62.89	26.02	23.43
Planctomycetes			2.78

1298

1299

1300 Table S5: List of the gene names from UniProt and or CAZY that were used in the total counts
 1301 for the metagenome and whole genome isolates.

Gene product name	Enzyme
Cellobiose phosphorylase	Cellubiosidase
Cellobiose 2-epimerase	Cellubiosidase
pepA/Cytosol aminopeptidase	Leucine aminopeptidase
Putative Cytosol aminopeptidase	Leucine aminopeptidase
Aminopeptidase	Leucine aminopeptidase
Putative lipoprotein aminopeptidase Lpql	Leucine aminopeptidase
Cytosol Aminopeptidase	Leucine aminopeptidase
Aminopeptidase N	Leucine aminopeptidase
Dipeptidyl aminopeptidase 4	Leucine aminopeptidase
D-aminopeptidase	Leucine aminopeptidase
Dipeptidyl aminopeptidase BIII	Leucine aminopeptidase
Dipeptidyl aminopeptidase BII	Leucine aminopeptidase
Dipeptidyl aminopeptidase BI	Leucine aminopeptidase
N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase	N-acetyl-B-D-glucosamididase
tagA	N-acetyl-B-D-glucosamididase
N-acetyltransferase domain-containing protein	N-acetyl-B-D-glucosamididase
Glycoside hydrolase superfamily	N-acetyl-B-D-glucosamididase
N-acetyl-beta-D glucosaminidase	N-acetyl-B-D-glucosamididase
Beta-N-acetylglucosaminidase	N-acetyl-B-D-glucosamididase
Nag	N-acetyl-B-D-glucosamididase
Phosphoserine phosphatase 1	Phosphatase
Phosphoglycolate phosphatase	Phosphatase
Alkaline phosphatase D	Phosphatase
Alkaline phosphatase	Phosphatase
Phosphoadenosine phosphosulfate reductase	Phosphatase
Phosphoserine phosphatase RsbU	Phosphatase
Phosphoserine phosphatase SerB2	Phosphatase
Endo-1,4-beta-xylanase A/xynA1	Xylosidase
Beta-xylosidase/xynB	Xylosidase
Bifunctional xylanase/deacetylase/xynD	Xylosidase
Xyloglucanase	Xylosidase
Beta-xylosidase	Xylosidase

Glucan 1,4-alpha-glucosidase SusB	Alpha glucosidase
Beta-glucosidase BoGH3A	Beta glucosidase
Thermostable beta-glucosidase B	Beta glucosidase
Beta-glucosidase A	Beta glucosidase
Beta-N-acetylglucosaminidase/beta-glucosidase	Beta glucosidase
Beta-glucosidase B	Beta glucosidase
Glucan endo-1,3-beta-glucosidase A1	Beta glucosidase
Beta-glucosidase BoGH3B	Beta glucosidase
Beta-glucosidase	Beta glucosidase
putative 6-phospho-beta-glucosidase	Beta glucosidase
Oligo-1,6-glucosidase	Beta glucosidase
cellobiose phosphorylase	Cellubiosidase
Cellobiose 2-epimerase	Cellubiosidase
pepA/Cytosol aminopeptidase	Leucine aminopeptidase
Putative Cytosol aminopeptidase	Leucine aminopeptidase

1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316

Table S6: Number of individual gene counts present for each MAG that had annotated genes for the seven enzymes: α -glucosidase (AG), β -glucosidase (BG), β -D-cellubiosidase (CB), leucine aminopeptidase (LAP), N-Acetyl- β -D-glucosaminidase (NAG), phosphatase (PHOS), and β -xylosidase (XYL).

Site and Sample Depth	Phyla	Class	AG	BG	CB	LAP	NAG	PHOS	XYL
BPF1-22-24_bin_2	Chloroflexota	Ellin6529	0	1	0	3	2	6	0
BPF1-28-30_bin_5	Proteobacteria	Gammaproteobacteria	0	1	1	1	2	3	1
BPF1-28-30_bin_8	Actinobacteriota	Thermoleophilia	0	0	0	1	2	5	0
BPF1-30-32_bin_3	Chloroflexota	Ellin6529	0	2	0	3	2	7	0
BPF1-32-34_bin_1	Chloroflexota	Ellin6529	0	1	0	3	2	7	0
BPF1-34-36_bin_3	Actinobacteriota	Thermoleophilia	0	1	0	1	2	5	0
BPF1-38-40_bin_4	Actinobacteriota	Thermoleophilia	0	0	0	1	2	8	0
BPF1-48-50_bin_4	Actinobacteriota	Thermoleophilia	0	1	0	1	2	5	0
BPF1-50-52_bin_1	Actinobacteriota	Thermoleophilia	0	1	0	1	2	5	0
BPF1-54-56_bin_1	Actinobacteriota	Thermoleophilia	0	1	0	1	2	4	0

BPF5-00-01_bin_1	Cyanobacteria	Cyanobacteriia	0	0	1	2	2	6	0
BPF1-00-02_bin_3	Planctomycetota	Phycisphaerae	0	0	0	0	1	1	0
BPF1-04-06_bin_3	Planctomycetota	Phycisphaerae	0	0	0	2	1	6	0
BPF1-04-06_bin_4	Acidobacteriota	Blastocatellia	0	0	0	14	1	6	1
BPF1-10-12_bin_1	Acidobacteriota	Blastocatellia	0	0	0	6	1	2	1
BPF1-10-12_bin_3	Chloroflexota	Ellin6529	0	1	0	2	1	6	0
BPF1-12-14_bin_1	Planctomycetota	Phycisphaerae	0	0	0	1	1	5	0
BPF1-14-16_bin_1	Chloroflexota	Ellin6529	0	1	0	3	1	6	0
BPF1-14-16_bin_2	Acidobacteriota	Blastocatellia	0	0	0	17	1	6	1
BPF1-16-18_bin_1	Chloroflexota	Ellin6529	0	1	0	3	1	6	0
BPF1-16-18_bin_4	Acidobacteriota	Blastocatellia	0	0	0	18	1	7	1
BPF1-16-18_bin_5	Proteobacteria	Gammaproteobacteria	0	2	0	2	1	1	1
BPF1-18-20_bin_2	Acidobacteriota	Blastocatellia	0	0	0	16	1	6	1
BPF1-18-20_bin_3	Acidobacteriota	Blastocatellia	0	0	0	16	1	4	0
BPF1-20-22_bin_2	Acidobacteriota	Blastocatellia	0	0	0	10	1	3	0
BPF1-20-22_bin_3	Acidobacteriota	Blastocatellia	0	0	0	17	1	5	1
BPF1-20-22_bin_4	Acidobacteriota	Vicinamibacteria	0	0	0	13	1	5	3
BPF1-22-24_bin_1	Acidobacteriota	Vicinamibacteria	0	0	0	9	1	2	1
BPF1-22-24_bin_3	Actinobacteriota	Thermoleophilia	0	0	0	1	1	5	0
BPF1-24-26_bin_3	Chloroflexota	Ellin6529	0	1	0	1	1	6	0
BPF1-24-26_bin_5	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-26-28_bin_2	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-32-34_bin_4	Actinobacteriota	Thermoleophilia	0	0	0	1	1	5	0
BPF1-34-36_bin_2	Planctomycetota	Planctomycetes	0	0	0	5	1	6	0
BPF1-34-36_bin_5	Acidobacteriota	Vicinamibacteria	0	0	0	15	1	5	3
BPF1-34-36_bin_7	Chloroflexota	Ellin6529	0	0	0	2	1	3	0
BPF1-36-38_bin_2	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-36-38_bin_6	Acidobacteriota	Blastocatellia	0	0	0	11	1	2	0
BPF1-38-40_bin_1	Actinobacteriota	Thermoleophilia	0	1	0	1	1	5	0
BPF1-38-40_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	1	1	2	0
BPF1-38-40_bin_5	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-40-42_bin_1	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-40-42_bin_2	Acidobacteriota	Blastocatellia	0	0	0	17	1	6	1
BPF1-42-44_bin_1	Acidobacteriota	Blastocatellia	0	0	0	16	1	6	1
BPF1-42-44_bin_2	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-44-46_bin_2	Acidobacteriota	Blastocatellia	0	0	0	10	1	6	1
BPF1-44-46_bin_3	Chloroflexota	Ellin6529	0	2	0	2	1	7	0
BPF1-44-46_bin_4	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-46-48_bin_1	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3

BPF1-48-50_bin_3	Acidobacteriota	Vicinamibacteria	0	0	0	14	1	5	3
BPF1-52-54_bin_4	Actinobacteriota	Thermoleophilia	0	0	0	1	1	5	0
BPF2-00-02_bin_1	Actinobacteriota	UBA4738	0	1	0	1	1	2	0
BPF2-00-02_bin_2	Acidobacteriota	Blastocatellia	0	0	0	20	1	7	4
BPF2-02-04_bin_1	Acidobacteriota	Blastocatellia	0	0	0	18	1	5	0
BPF2-04-06_bin_1	Acidobacteriota	Blastocatellia	0	0	0	22	1	4	1
BPF2-06-08_bin_3	Acidobacteriota	Blastocatellia	0	0	0	18	1	6	1
BPF2-06-08_bin_5	Acidobacteriota	Mor1	0	0	0	2	1	3	0
BPF2-08-10_bin_1	Acidobacteriota	Blastocatellia	0	0	0	23	1	4	1
BPF2-12-14_bin_2	Acidobacteriota	Blastocatellia	0	0	0	24	1	4	1
BPF2-14-16_bin_1	Acidobacteriota	Blastocatellia	0	1	0	25	1	4	1
BPF2-16-18_bin_1	Acidobacteriota	Blastocatellia	0	0	0	24	1	4	1
BPF2-20-22_bin_2	Acidobacteriota	Blastocatellia	0	0	0	26	1	4	1
BPF2-22-24_bin_2	Acidobacteriota	Blastocatellia	0	0	0	24	1	4	0
BPF2-24-26_bin_1	Chloroflexota	Ellin6529	0	0	0	3	1	4	0
BPF2-24-26_bin_2	Acidobacteriota	Blastocatellia	0	0	0	19	1	4	1
BPF2-26-28_bin_1	Acidobacteriota	Blastocatellia	0	1	0	14	1	3	1
BPF2-28-30_bin_1	Acidobacteriota	Blastocatellia	0	0	0	18	1	4	1
BPF3-00-01_bin_2	Chloroflexota	Ellin6529	0	1	0	8	1	5	0
BPF3-01-02_bin_2	Chloroflexota	Ellin6529	0	1	0	3	1	3	0
BPF3-02-03_bin_3	Chloroflexota	Ellin6529	0	1	0	6	1	1	0
BPF3-02-03_bin_4	Actinobacteriota	UBA4738	0	0	0	2	1	3	0
BPF1-00-02_bin_1	Acidobacteriota	Thermoanaerobaculia	0	1	0	13	0	7	1
BPF1-00-02_bin_2	Bacteroidota	Bacteroidia	0	0	0	3	0	0	0
BPF1-02-04_bin_1	Acidobacteriota	Thermoanaerobaculia	0	1	0	14	0	7	0
BPF1-02-04_bin_2	Planctomycetota	Planctomycetes	0	0	0	7	0	6	0
BPF1-04-06_bin_1	Verrucomicrobiota	Verrucomicrobiae	0	0	0	0	0	1	0
BPF1-04-06_bin_2	Proteobacteria	Alphaproteobacteria	0	0	0	2	0	0	0
BPF1-06-08_bin_1	Acidobacteriota	Thermoanaerobaculia	0	1	0	15	0	7	1
BPF1-06-08_bin_2	Proteobacteria	Alphaproteobacteria	0	0	0	2	0	0	0
BPF1-08-10_bin_1	Acidobacteriota	Thermoanaerobaculia	0	1	0	10	0	2	0
BPF1-08-10_bin_2	Proteobacteria	Alphaproteobacteria	0	0	0	2	0	1	0
BPF1-08-10_bin_3	Actinobacteriota	Thermoleophilia	0	0	0	1	0	4	0
BPF1-08-10_bin_4	Acidobacteriota	Blastocatellia	0	0	0	0	0	4	0
BPF1-10-12_bin_2	Acidobacteriota	Thermoanaerobaculia	0	1	0	15	0	7	1
BPF1-14-16_bin_3	Verrucomicrobiota	Verrucomicrobiae	0	0	0	2	0	1	0
BPF1-16-18_bin_2	Gemmatimonadota	Gemmatimonadetes	1	2	0	5	0	2	0
BPF1-16-18_bin_3	Acidobacteriota	Blastocatellia	0	0	0	2	0	2	0
BPF1-18-20_bin_1	Actinobacteriota	Thermoleophilia	0	0	0	2	0	6	0

BPF1-20-22_bin_1	Actinobacteriota	Thermoleophilia	0	0	0	2	0	7	0
BPF1-22-24_bin_4	Acidobacteriota	Acidobacteriae	0	0	1	3	0	0	0
BPF1-24-26_bin_1	Actinobacteriota	Thermoleophilia	0	1	0	0	0	4	0
BPF1-24-26_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	2	0	6	0
BPF1-24-26_bin_4	Proteobacteria	Gammaproteobacteria	0	0	0	2	0	1	0
BPF1-26-28_bin_1	Actinobacteriota	Thermoleophilia	0	0	0	2	0	7	0
BPF1-28-30_bin_1	Verrucomicrobiota	Verrucomicrobiae	0	0	0	1	0	0	0
BPF1-28-30_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	2	0	6	0
BPF1-28-30_bin_3	Proteobacteria	Gammaproteobacteria	0	0	0	3	0	4	0
BPF1-28-30_bin_4	Acidobacteriota	Blastocatellia	0	0	0	1	0	3	0
BPF1-28-30_bin_6	Chloroflexota	Ellin6529	0	1	0	3	0	6	0
BPF1-28-30_bin_7	Actinobacteriota	Thermoleophilia	0	2	0	4	0	4	0
BPF1-30-32_bin_1	Proteobacteria	Gammaproteobacteria	0	0	0	3	0	4	0
BPF1-30-32_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	1	0	3	0
BPF1-30-32_bin_4	Chloroflexota	Ellin6529	0	2	0	0	0	1	0
BPF1-30-32_bin_5	Actinobacteriota	Thermoleophilia	0	1	0	3	0	5	0
BPF1-30-32_bin_6	Actinobacteriota	Thermoleophilia	0	0	0	2	0	7	0
BPF1-32-34_bin_2	Actinobacteriota	Thermoleophilia	0	2	0	4	0	4	0
BPF1-32-34_bin_3	Proteobacteria	Gammaproteobacteria	0	0	0	3	0	4	0
BPF1-32-34_bin_5	Acidobacteriota	Blastocatellia	0	1	0	3	0	4	0
BPF1-34-36_bin_1	Chloroflexota	Ellin6529	0	2	0	2	0	4	0
BPF1-34-36_bin_4	Proteobacteria	Gammaproteobacteria	0	0	0	3	0	4	0
BPF1-34-36_bin_6	Actinobacteriota	Thermoleophilia	0	0	0	1	0	4	0
BPF1-34-36_bin_8	Actinobacteriota	Thermoleophilia	0	0	0	2	0	6	0
BPF1-36-38_bin_1	Actinobacteriota	Thermoleophilia	0	1	0	1	0	6	0
BPF1-36-38_bin_3	Proteobacteria	Gammaproteobacteria	0	0	0	2	0	2	0
BPF1-36-38_bin_5	Actinobacteriota	Thermoleophilia	0	0	0	0	0	2	0
BPF1-36-38_bin_7	Actinobacteriota	Thermoleophilia	0	1	0	4	0	4	1
BPF1-38-40_bin_3	Actinobacteriota	Thermoleophilia	0	1	0	4	0	4	1
BPF1-44-46_bin_1	Gemmatimonadota	Gemmatimonadetes	1	1	0	10	0	3	1
BPF1-44-46_bin_5	Acidobacteriota	Blastocatellia	0	0	1	2	0	0	0
BPF1-46-48_bin_2	Gemmatimonadota	Gemmatimonadetes	0	2	0	6	0	4	0
BPF1-46-48_bin_3	Chloroflexota	Ellin6529	0	1	0	1	0	2	0
BPF1-48-50_bin_1	Chloroflexota	Ellin6529	0	3	0	2	0	3	0
BPF1-48-50_bin_2	Actinobacteriota	Acidimicrobiia	0	0	0	1	0	1	0
BPF1-50-52_bin_2	Proteobacteria	Gammaproteobacteria	0	0	0	1	0	4	0
BPF1-52-54_bin_1	Acidobacteriota	Vicinamibacteria	0	0	0	4	0	4	0
BPF1-52-54_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	2	0	6	0
BPF1-52-54_bin_3	Chloroflexota	Ellin6529	0	3	0	2	0	2	0

BPF1-54-56_bin_2	Methylomirabilota	Methylomirabilia	0	0	0	4	0	7	0
BPF1-56-58_bin_1	Actinobacteriota	Thermoleophila	0	0	0	1	0	4	0
BPF1-56-58_bin_2	Eisenbacteria	RBG-16-71-46	0	1	1	3	0	2	0
BPF1-56-58_bin_3	Actinobacteriota	Acidimicrobiia	0	1	0	3	0	5	0
BPF1-56-58_bin_4	Actinobacteriota	Thermoleophila	0	1	0	1	0	4	0
BPF1-56-58_bin_5	Chloroflexota	Ellin6529	0	3	0	2	0	4	0
BPF2-06-08_bin_1	Acidobacteriota	Blastocatellia	0	1	0	13	0	5	3
BPF2-06-08_bin_2	Proteobacteria	Gammaproteobacteria	0	0	0	2	0	3	0
BPF2-06-08_bin_4	Acidobacteriota	Thermoanaerobaculia	0	0	0	3	0	1	0
BPF2-08-10_bin_2	Acidobacteriota	Blastocatellia	0	0	0	7	0	3	0
BPF2-08-10_bin_3	Actinobacteriota	Actinomycetia	0	2	0	4	0	4	0
BPF2-08-10_bin_4	Chloroflexota	Anaerolineae	1	3	0	1	0	1	0
BPF2-12-14_bin_1	Acidobacteriota	Blastocatellia	0	1	0	7	0	2	0
BPF2-14-16_bin_2	Actinobacteriota	Thermoleophila	0	1	0	2	0	4	0
BPF2-14-16_bin_3	Acidobacteriota	Blastocatellia	0	1	1	3	0	2	0
BPF2-20-22_bin_1	Actinobacteriota	Thermoleophila	0	0	0	2	0	6	0
BPF2-22-24_bin_1	Actinobacteriota	Thermoleophila	0	1	0	2	0	6	0
BPF2-24-26_bin_3	Actinobacteriota	Thermoleophila	0	0	0	1	0	3	0
BPF2-24-26_bin_4	Actinobacteriota	Thermoleophila	0	0	0	1	0	4	0
BPF3-00-01_bin_1	Actinobacteriota	Thermoleophila	0	0	0	3	0	6	0
BPF3-00-01_bin_3	Actinobacteriota	Acidimicrobiia	0	1	0	3	0	7	0
BPF3-01-02_bin_1	Actinobacteriota	Thermoleophila	0	0	0	3	0	7	0
BPF3-02-03_bin_1	Proteobacteria	Alphaproteobacteria	0	0	0	5	0	6	0
BPF3-02-03_bin_2	Actinobacteriota	Thermoleophila	0	0	0	1	0	4	0
BPF3-18-20_bin_1	Actinobacteriota	Thermoleophila	0	1	0	3	0	7	0
BPF4-00-01_bin_1	Chloroflexota	Ellin6529	0	1	0	3	0	2	0
BPF4-00-01_bin_2	Proteobacteria	Alphaproteobacteria	0	0	0	7	0	0	0
BPF4-02-03_bin_1	Acidobacteriota	Blastocatellia	0	1	0	9	0	2	3
BPF4-02-03_bin_2	Acidobacteriota	Thermoanaerobaculia	0	1	0	5	0	2	0
BPF4-02-03_bin_3	Actinobacteriota	Actinomycetia	0	0	0	4	0	8	0
BPF4-43-45_bin_1	Acidobacteriota	Thermoanaerobaculia	0	0	1	7	0	1	0
BPF4-43-45_bin_2	Proteobacteria	Gammaproteobacteria	0	0	0	2	0	3	0
BPF4-43-45_bin_3	Actinobacteriota	Thermoleophila	0	0	0	1	0	2	0
BPF4-43-45_bin_4	Chloroflexota	Anaerolineae	0	6	0	5	0	5	0
BPF4-43-45_bin_5	Chloroflexota	Dehalococcoidia	0	0	0	1	0	1	0
BPF5-00-01_bin_2	Actinobacteriota	Thermoleophila	0	0	0	1	0	1	0
BPF5-01-02_bin_1	Acidobacteriota	Blastocatellia	0	1	0	12	0	2	1
BPF5-01-02_bin_2	Actinobacteriota	Thermoleophila	0	0	0	1	0	2	1
BPF5-02-03_bin_1	Acidobacteriota	Blastocatellia	0	1	0	14	0	2	1

BPF5-02-03_bin_2	Actinobacteriota	Thermoleophilia	0	0	0	1	0	5	0
BPF1-36-38_bin_4	Thermoproteota	Nitrososphaeria	0	0	0	1	0	2	0
BPF2-24-26_bin_5	Thermoproteota	Nitrososphaeria	0	0	0	0	0	2	0

1317
1318