Microbial life in ancient permafrost along a salinity gradient illuminated by metagenomics

Xiaofen Wu  
University of Tennessee

Abraham L. Almatari  
University of Tennessee

Daniel E. Williams  
University of Tennessee

Susan M. Pfiffner  
University of Tennessee

Elizaveta M. Rivkina  
Russian Academy of Sciences

Karen G. Lloyd  
University of Tennessee

Tatiana A. Vishnivetskaya (tvishniv@utk.edu)  
University of Tennessee

Research Article

Keywords: Arctic, permafrost, metagenomics, metagenome-assembled genomes, metabolic potential.

Posted Date: September 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1954178/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

This study describes the composition and potential metabolic adaptation of microbial communities in northeastern Siberia, a repository of the oldest permafrost in the Northern Hemisphere. Samples of contrasting depth (1.75 to 25.1 m below surface) and age (from ~10 kyr to 1.1 Myr) were collected from freshwater permafrost (FP) of borehole AL1_15 on the Alazeya River, and freshwater coastal permafrost (FCP) overlying marine permafrost (MP) of borehole CH1_17 on the East Siberian Sea coast. Using 16S rRNA gene sequencing, we showed that the biodiversity decreased dramatically with permafrost age and separated the samples into three groups: FP and FCP together (10–100 kyr old), FP (>900 kyr old), and MP (105–120 kyr old). Younger FP/FCP deposits were distinguished by the presence of Acidobacteriota, Bacteroidota, Chloroflexota_A, and Gemmatimonadota, older FP deposits had a higher proportion of Gammaproteobacteria, and older MP deposits had much more uncultured groups within Asgardarchaeota, Crenarchaeota, Chloroflexota, Patescibacteria, and unassigned archaea. The 60 recovered metagenome-assembled genomes (MAGs) and un-binned metagenomic assemblies suggested that despite the large taxonomic differences between samples, they all had a wide range of taxa capable of fermentation coupled to nitrate utilization, with the exception of sulfur reduction present only in old MP deposits.

Introduction

Permafrost is a unique perennially frozen environment that covers ~24% of the land surface area in Earth's Northern Hemisphere [1]. Permafrost temperature varies from around −2°C in Svalbard [2] and near the southern permafrost boundary in Siberia to the lower temperature of −18°C on Ellesmere Island, Canada [3]. Permafrost deposits are heterogenous and their physicochemical properties are determined by their origin (e.g., lacustrine and alluvial versus marine) and formation (e.g., epigenetic or consecutively frozen versus syngenetic or simultaneously frozen) [4]. Permafrost temperature and size of mineral particles (e.g., clay, silt or sand) affect the amount and thickness of unfrozen water films, which habitually accounts for 1.5-7% of total water [5], yet may form thicker films and brine lenses (cryopegs) with free supercooled saline water in mineralization up to 250 g L⁻¹ and low water activity [6]. Arctic permafrost is a major reservoir for stored organic carbon [7–9], however frozen sequestered carbon is not very available as a nutritional substrate within the frozen ground.

Despite the freezing temperatures, insignificant content of liquid water, low water activity, and nutrient deprivation, viable and at times metabolically active microorganisms are confirmed to be present in permafrost. Several lines of evidence support this indication, including (i) isolation of viable culturable cells [10]; (ii) observation of viable non-culturables using LIVE/DEAD differential staining coupled with fluorescence microscopy [11, 12]; (iii) radioactively labeled isotope-probing evidence of biomass production and genome replication of permafrost microorganisms at subzero temperatures [13, 14]; (iv) observation of viable and metabolically active cells using aspartic acid racemization [15]; (v) presence of metabolic pathways involved in microbial adaptations to permafrost conditions [16]; and (vi) implication of metabolic activity and cell division based on metatranscriptomes and metaproteomes in intact
permafrost [17]. A metagenomic study of old Siberian permafrost samples showed that microbial community composition depends on the conditions when permafrost deposits formed. The late Pleistocene ice-rich syngenetically frozen sediments accumulated under cold arid aerobic conditions showed absence of detectable methane and a low abundance of methanogenic archaea and genes encoding carbon and nitrogen related functions but a higher abundance of genes associated with sulfur cycling [18]. The opposite data were shown for the late Pleistocene floodplain lake-alluvial sediments, which had formed epigenetically in anaerobic conditions and exhibited a presence of 1.2 mmol kg$^{-1}$ biogenic methane and up to 1.5% of methanogenic Archaea [18]. The existence of an indigenous microbial communities in cryopegs encapsulated in ancient epigenetically frozen marine sediments (up to 200 kyr) was demonstrated by cultivation studies [19]. Permafrost microbes surrounded by microscopic layers of liquid brines may develop adaptations to this environment as shown through comparative analyses of metagenome-assembled genomes (MAGs) derived from syngenetic permafrost with genomes of non-permafrost counterparts [16].

The perennially frozen deposits in the Northern Hemisphere store large amounts of soil organic carbon [8]. The preserved soil organic carbon is potentially vulnerable to remobilization and easily becomes accessible to microbial degradation following the permafrost thawing as a result of increase in global temperatures [20]. Increases in salinity and consecutive decrease in water potential of permafrost deposits will cause a formation of specialized microbial communities due to changes in population structure and cellular adaptation in order to withstand salinization [21, 22]. Salt adapted microbial communities may affect the emission of greenhouse gasses such as carbon dioxide (CO$_2$), methane (CH$_4$), and nitrous oxide (N$_2$O) [23]. Study of agricultural soils of different salinities showed that high salinity decreases bacterial phylogenetic diversity, destabilizes bacterial networks and inhibits key nitrogen metabolisms [24]. However, the impact of salinity on permafrost bacterial community composition and community-level function remains uncertain. The current dataset provides a unique opportunity to compare the long-term impact of low temperature and varying salinity on permafrost microbial community. Characterization of microbial communities and their metabolic potential from diverse permafrost deposits becomes of ultimate importance due to global warming and increasing permafrost thawing. The latter process will result in release of frozen microbes and unleashing of microbial processes that in turn will contribute to the decomposition of sequestered organic carbon, global flux of greenhouse gasses and other consequences, for example, release of mercury from thawing permafrost [25].

In this research we studied permafrost samples collected from epigenetically formed permafrost strata in northeastern Siberia. The permafrost in that area developed during cold periods in the Late Pliocene, Pleistocene and Holocene epochs and extends to hundreds of meters in depth forming layers of different ages and different origin, e.g., freshwater, freshwater coastal and marine [4, 26]. The deep layers represent one of the oldest continuously frozen localities on Earth [6]. Therefore, permafrost deposits of northeastern Siberia that did not thaw during the Holocene climatic optimum are of great interest for microbial community analyses. Much of our knowledge about Siberian permafrost is derived from
cultivation studies, which revealed communities of anaerobic and aerobic, spore-forming and non-spore-forming bacteria, green algae, yeast, fungi as well as highly specialized organisms like methanogenic archaea [10, 27, 28]. Unlike culture-dependent methods, which uncover limited diversity, and 16S rRNA gene sequencing, which identifies the taxonomic composition of bacterial and archaeal communities, metagenomic sequencing allows the structural characterization of the whole microbial communities and their metabolic potential.

We used both 16S rRNA gene amplicon and metagenome sequencing to investigate the taxonomic composition, genetic diversity and metabolic potential of permafrost microbial communities in samples of epigenetically frozen sediments along a salinity gradient. The northeastern Siberia permafrost samples were from different depths (1.75–25.1 m), ages (~ 10 kyr-1.1 Myr) and salinity types (freshwater, freshwater coastal and marine permafrost). The metagenome data were assembled and partitioned into MAGs to reconstruct metabolic pathways for the dominant members of the microbial communities. To complement the genome-based metagenomics, a gene-centric analysis was performed by direct examination of un-binned and assembled metagenomic data.

Materials And Methods

1.1 Sampling Sites and Collection of Samples

The sampling sites were located within the Kolyma Lowland in northeastern Siberia (Fig. S1). This area is situated in an intermontane depression containing wide river valleys, maritime plains and seacoasts [29]. The study sites were selected to access epigenetic (frozen after sedimentation) permafrost of freshwater and marine origin [4]. The first site (AL1_15) was located on the right bank of the Alazeya River (N69º20.438, E154º59.713); the second site (CH1_17) was located at a distance of ~225 km north-east on the Chukochy Cape at the East Siberian Sea coast (N70º04.903, E159º52.282). The permafrost cores (25.8 m from borehole AL1_15 and 21.7 m from borehole CH1_17) were unearthed using UKB 12/25 drilling rig (Machine-Building Plant named after V.V. Vorovsky, Yekaterinburg) that operates without any drilling fluids. Following a protocol for aseptic sample collection described previously [15], the outside layer of extracted cores was trimmed away with sterile knife and the permafrost samples were placed into sterile Whirl-Pak® sample bags and stored at -20°C until analyses. Temperature inside boreholes was measured at different depths with the HOBO U12-008 data loggers using Air/Water/Soil Temperature Sensors (Onset Computer Corporation, Bourne, MA).

1.2 Sample Description and Geochemical Analysis

Samples collected from borehole AL1_15 at depths of 2.95, 6.0, 15.0 and 25.1 m below the surface represented perennially frozen alluvial-lacustrine loams and silt loams defined as freshwater permafrost (FP) (Table S1). The sediments uncovered by borehole CH1_17, based on the physicochemical and micropaleontological analyses, were formed in a coastal zone with a changing water regime [30].
Therefore, the silt loams from the upper horizon (1.75-5.4 m) of borehole CH1_17 were identified to be freshwater coastal permafrost (defined as FCP) and deeper layers (depths from 11.5 to 19.6 m) were recognized to be saline silt loams and sandy loams of marine permafrost (defined as MP) (Table S1). Detailed geochemical characteristics were described previously [30, 31]. The salinity was calculated from conductivity using the equation described earlier [32]. Total carbon and nitrogen were determined on finely grounded permafrost samples at the Stable Isotope Laboratory at the University of Tennessee, Knoxville, United States. Water extract analysis, concentrations of methane and sulfate were obtained at the Soil Cryology Laboratory, Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Russia. The isotopic signature of methane was done using a custom built near-IR Continuous Wave Cavity Ring-Down Spectrometer at the Princeton University, Princeton, United States.

1.3 DNA Extraction and Sequencing

The total community genomic DNA (gDNA) was isolated in duplicate from four (2.95, 6.0, 15.0 and 25.1 m) samples of borehole AL1_15 and five (1.75, 5.4, 11.5, 16.9 and 19.6 m) samples of borehole CH1_17. The gDNA was extracted from ~0.5 g of soil using FastDNA Spin Kit for Soil according to the manufacturer's instructions (MP Biomedicals). Due to the low biomass, eight to sixteen extractions were pooled for each replicate sample. The pooled gDNA from each sample was concentrated using 100% ethanol (2-2.5 times of the pooled DNA volume) and 5 M NaCl (0.08 times). The mixture was centrifuged at 14,000 g for 25 min at room temperature. The obtained gDNA pellet was resuspended with 35 to 50 µl of DES water (from the FastDNA Spin Kit) and quantified using Qubit 3.0 Fluorometer (Life Technologies). Amount of gDNA varied from 4.8 to 540 ng per 1 g of permafrost.

The V4 region of archaeal and bacterial 16S rRNA genes were amplified using primer set 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GGACTACNVGGGTWTCTAAT) [33]. Samples with higher gDNA concentration were diluted 1:10, and 2.5 ml of gDNA was taken resulting in 1-5 ng gDNA per 25 ml reaction. After library preparation, the samples were normalized and pooled to a final concentration of 4 pM, combined with 10% PhiX, and loaded on a v3, 600 cycle flow cell, reading 2 x 275 bp paired-end on the Illumina MiSeq. A blank sample was used for sequencing control. Library preparation and sequencing were carried out at Genomics Core Facility at University of Tennessee, Knoxville, USA.

The same DNA extracts were used for metagenome libraries preparation using the Nextera DNA Flex Library Preparation Kit and sequenced on Illumina NovaSeq 6000 platform (2’ 150 bp) at Genomics Core Facility at Princeton University, USA. The input for sequencing was 1-20 ng DNA. The quality metagenomic sequencing data were obtained for four samples of borehole AL1_15 and three samples from borehole CH1_17.

1.4 Bioinformatic Analysis
The bioinformatic analyses were carried out using the Infrastructure for Scientific Applications and Advanced Computing resources at the University of Tennessee (project number UTK0011). From amplicon sequencing, a total number of 2.68 million sequences, ranging from 77,598 to 238,405 sequences per sample, were obtained and analyzed using DADA2 v1.10.0 [34]. Briefly, raw reads were filtered and trimmed to remove low quality reads with filtering parameters: truncLen=c(250, 200), maxN=0, maxEE=c(2, 2), truncQ=2, and trimLeft=c(19, 20). Trimmed paired-end reads were then merged and used to construct an amplicon sequence variant (ASV) table. ASVs which were found in the sequencing control were removed from the samples using Decontam v1.13.0 [35]. The identified ASVs were annotated using the GTDB database release 89 [36] after chimeric sequences were removed from the ASV table. Taxonomic composition analysis was performed using Phyloseq v1.32.0 [37] and visualized using ggplot2 v3.3.5 [38]. For further analysis, the ASV count was normalized with Phyloseq into percentage as the relative abundance. The alpha diversity indices were plotted using plot_richness function in Phyloseq by choosing measures of “Shannon (H)” and “Simpson (D)”.

Metagenome analysis was carried out first by checking qualities of the raw reads using fastqc. Adapters and low-quality reads were then removed with Trimmomatic v0.36 using default parameters [39]. Trimmed reads from all the metagenomes were co-assembled into contigs using Megahit v1.2.9 [40] with a kmer list of 31, 41, 51, 61, 71 and 81. The trimmed reads from each sample were then mapped back to the co-assembled contigs using bowtie2 v2.2.9 [41], samtools v1.3.1 [42] and Picard Markduplicates [43] with default parameters. Binning was carried out using Metabat v2.12.1 [44] with contig length cutoff of 1500 bp to generate near complete genomes. The quality of all the generated bins was evaluated using CheckM [45]. Bins with a completeness ³ 50% and contamination £ 10% were selected (defined as metagenome assembled genomes (MAGs)). Taxonomic annotation was performed on selected MAGs using GTDB-Tk v1.0.1 against GTDB database release 89 [36]. Prodigal v2.6.3 [46] was used to predict genes for each selected MAG, followed by annotation with kofamscan v1.2.0 against KEGG database [47]. Coverage for each MAG was obtained using CoverM after removing the 5% of bases with highest and lowest coverages (https://github.com/wwood/CoverM). Relative abundance was calculated by dividing coverage from each MAG by the total coverage from the corresponding sample. Mean relative abundance was used for samples with technical replicates (Table S2). A network analysis between MAGs and the samples were performed using Cytoscape v3.7.2 [48]. A neighbor-joining phylogenetic tree of cbbL and nifH gene sequences were constructed, respectively, using extracted amino acid sequences and gene reference databases [49, 50]. The sequences were first aligned using MAFFT version 7 [51], and the closely related sequences were visualized using Archaeopteryx v0.9921 [52].

Co-assembled contigs (³ 1000 bp) were annotated with Kofamscan v1.2.0 against KEGG database [47] after gene prediction with prodigal v2.6.3 [53]. Transcripts per million (TPM) was used as gene abundance and calculated using Salmon v1.0.0 [54]. TPM, which is normally used in RNA sequencing, represents reads per million when it is applied to metagenomes. Heatmaps were generated using R packages ComplexHeatmap v2.5.5 [55] and Pheatmap v1.0.12 [56]. Mean normalized TPM abundance was calculated for samples with technical replicates (Figures S3 and S5).
Results And Discussion

2.1 Geochemical Characteristics

Permafrost in the Alazeya River basin from borehole AL1_15 upper layers (FP 2.95-6.0 m) was formed ~10-40 kyr ago, while deeper permafrost sediments (FP 15.0 and 25.1 m) were formed ~0.9 and 1.1 Myr ago (Table S1) [4]. The geological age of permafrost sediments from borehole CH1_17 also increased with depth. The permafrost layer from FCP 1.75-5.8 m likely formed ~10-100 kyr ago, whereas permafrost from MP 11.5-19.6 m was formed about ~105-120 kyr ago (Table S1) [4, 6]. Temperature in borehole AL1_15 varied from -1.8°C in upper layers to -5.9°C in deeper layers, whereas temperatures in borehole CH1_17 were lower and decreased with depth from -4.8°C in FCP sediments to -8.1°C in deeper MP layers. Redox potential measured in MP sediments was +100 – (+150) mV [6] and was lower in FP varying from +40 – (–100) mV in upper layers to –256 mV in deeper layers [27] indicating anaerobic conditions in permafrost.

FP samples from AL1_15 had 0.04-0.11% total dissolved solids and low salinity in the range of 0.1-0.2 ppt (parts per thousand). FCP from CH1_17 formed in a coastal zone had 0.09-0.38% total dissolved solids and salinity 0.3-1.3 ppt increasing with depth. The concentration of total dissolved solids and salinity increased to 1.58% and 6.1 ppt, respectively (Table S1), in the deeper MP layers in the same borehole. Concentration of ions Cl\(^-\) and Na\(^+\) showed a similar trend in both boreholes [15, 30], increasing with depth from 7.5 and 0.89 mmol kg\(^{-1}\) to 25 and 1.76 mmol kg\(^{-1}\), respectively, in freshwater AL1_15 sediments; and from 6.5-55 and 3.7-33.2 mmol kg\(^{-1}\) in freshwater coastal CH1_17 sediments to 80-230 and 68-193.6 mmol kg\(^{-1}\), respectively, in marine permafrost CH1_17 sediments (Table S1). Concentration of K\(^+\) was in the range of 0.27-0.66 mmol kg\(^{-1}\) in FP; 0.3-0.6 mmol kg\(^{-1}\) in FCP; and increased from 1.8 to 7.1 mmol kg\(^{-1}\) with depth in MP (Table S1). Ratio of SO\(_4^{2-}\) to Cl\(^-\) in all samples was below 1 indicating the dominance of Cl\(^-\) anions in all studied permafrost samples. Cl\(^-\), Na\(^+\) and K\(^+\) are major osmotically active ions. In addition, Cl\(^-\) is widespread in bacteria and may be involved in the stabilization of membrane potential, regulation of intracellular pH gradients, regulation of key enzymes and salt adaptation [57].

The isotopic signature of methane collected from borehole AL1_15 at a depth of 23 m was -84.9 ± 0.5‰ for d\(^{13}\)C\(_{CH4}\) and -316.2 ± 1.5‰ for d\(^2\)H\(_{CH4}\) that shows a biogenic origin.

Total carbon in FP ranged from 0.762% to 2.663%; in FCP (1.678-2.088%); and in MP (0.214-0.911%). Total nitrogen measured in FP (0.072-0.175%); in FCP (0.149-0.158%); in MP at depth of 11.5-16.9 m (0.022-0.077%) and at 19.6 m (0.108%) was lower than total nitrogen in Siberian tundra soils (0.5-2%) [58]. During the previous study of freshwater permafrost samples collected from the Alazeya River site [59], a low concentration of ammonium (<100 ppm g\(^{-1}\) wet soil) and traces of nitrite and nitrate were detected in the upper sediments followed by a layer (12-24 m), where the concentration of ammonium was 3 times as high as that of the top layer. The same study [59] found high concentrations of nitrite (5-110 ppm g\(^{-1}\) wet soil), nitrate (65-115 ppm g\(^{-1}\) wet soil) and ammonium (110-152 ppm g\(^{-1}\) wet soil) in the
Chukochy Cape marine sediments formed 100-150 kyr, while only ammonium was detected in freshwater coastal sediments in concentration of 22-85 ppm g\(^{-1}\) wet soil.

### 2.2 Microbial Diversity From 16S rRNA Gene Sequencing

In the freshwater permafrost from AL1_15, alpha-diversity determined by the Shannon and Simpson indexes was higher in the shallow and younger permafrost samples and biodiversity decreased with age (Fig. 1 A). A similar trend was observed in CH1_17 (Fig. 1 A).

The decrease in diversity with age is in accordance with previous findings in ancient permafrost [15, 60]. Notably, despite the similar origin and close estimated age, biodiversity in MP_11.5 was higher than in MP_16.9 and MP_19.6 (Fig. 1 A), which may be related to conditions of sedimentation or small-scale heterogeneity connected to sample depth. The non-metric multidimensional scaling (NMDS) analysis of all amplicon sequence variants (ASVs, based on relative abundance) showed that the microbial communities differed between the sediments based on age and salinity (Fig. 1 B). In both boreholes the shallow and younger samples grouped separately from deeper and older permafrost. Despite having different salinities ranging from 0.1 ppt in young FP to 0.3-1.3 ppt in FCP, the FP that formed \(~10-40\) kyr ago and FCP that formed \(~10-40\) and \(100\) kyr grouped in close proximity to each other, separately from deeper, older, and more saline permafrost. Based on 16S rRNA gene relative amplicon abundance from duplicate extractions, Actinobacteriota was the most abundant phylum at 2.95, 6.0 and 15 m in FP, while microbial communities from 25.1 m were dominated by Gammaproteobacteria (Fig. 2).

The increase of Gammaproteobacteria in the deeper FP samples (15 and 25.1 m) could be explained by an increase in salinity. Previous studies of freshwater and saline lakes and ponds showed a similar trend when Actinobacteria were more abundant in freshwater and low salinity waters [61] while Gammaproteobacteria increased with increasing salinity [61, 62]. There was no correlation between abundance of Gammaproteobacteria and salinity in samples from CH1_17.

In CH1_17, the low-salinity top permafrost layers (1.75, 5.4 m) and 11.5 m were dominated by Actinobacteriota and shifted to Chloroflexota in MP at 16.9 and 19.6 m, while 11.5 m sample contained considerable amount of Chloroflexota along with Actinobacteriota. Members of the Chloroflexi phylum are dominant in numerous subseafloor environments [63] that is consistent with our finding of high amounts of Chloroflexota in marine permafrost. ASVs from spore-forming Clostridia (Firmicutes_A phylum) and Bacilli (Firmicutes phylum) classes were, on average, most abundant at 15.0 m in FP (7.41%). These ASVs increased from a mean relative abundance of 2.74% at 1.75 m to 20.69% at 5.4 m in FCP, and then declined to 2.10, 1.18 and 1.57% at 11.5, 16.9 and 19.6 m, respectively, in MP. The highest abundance of the Firmicutes was observed in FCP at 5.4 m with moderate salinity of 1.3 ppt (1.1 mS/cm) and methane concentration of 76 µmol/kg. Previous studies showed a positive correlation between relative abundances of Firmicutes and either salinity gradient of 0.36-6.72 mS/cm in temperate soils [64] or CH\(_4\) production in anaerobic digester [65]. Archaea were detected in all the samples except for FP_25.1.
Halobacterota and Crenarchaeota contributed a relatively high abundance in FCP at 5.4 m and in MP at 11.5 m (averaging 16.32 and 16.46%), respectively. Within the Halobacterota, ASVs associated with methanogens, such as *Methanobacteria*, *Methanosarcinia*, *Methanomicrobia*, *Syntrophoarchaeia*, and *Methanocellia*, were most abundant at 6.0 m in FP (averaging 2.09%) and at 5.4 m in FCP (averaging 8.52%). While these samples originated in lake and coastal mixing zones, respectively, and were collected at similar depths, the sample from 5.4 m had 10 times higher salinity, and more than twice the abundance of carbon, nitrogen, and methane (Table S1). The presence of methane at 5.4 m is supported by presence of salt-tolerant methanogens, such as *Methanobacteria*, *Methanocellia*, *Methanomicrobia* [66]. Decrease in abundance of the methanogens in deeper MP samples may be connected to the increase of SO$_4^{2-}$. Similar to our study, negative correlation between presence of SO$_4^{2-}$ and CH$_4$ production in artificial sea water microcosms was reported [23]. Methanogenic archaea were not identified at 15.0 and 25.1 m in FP despite a high concentration of methane in those samples. This methane could therefore be explained by methane compression in deeper layers during epigenetic freezing from the top down [67].

### 2.3 Identification and Distribution of MAGs

Metagenomic binning from the co-assembly of the seven permafrost samples (two marine permafrost samples failed to yield metagenomes of good quality) resulted in a total of 60 MAGs with $\geq 50\%$ completeness and $\leq 10\%$ contamination, altogether contributing 0.24 to 59.10% of sequencing reads of the respective samples, with the highest proportion of reads in FP_25.1 and lowest in FP_2.95 (Fig. 3 A). The low percentage of reads recovered in MAGs from young permafrost samples may be attributed to higher biodiversity in that sample (Fig. 1 A). Because of different coverages of metagenomes, the co-assembly provided obvious benefits for capturing more of the diversity due to higher read depth, robust assembly, improved MAGs recovery [68], and facilitated comparison across permafrost samples. The reconstructed MAGs comprised phylogenetically diverse members from 2 archaeal and 15 bacterial phyla (Fig. 3 A and Table S2). Of the 60 recovered MAGs, 15 MAGs were retrieved from both boreholes, while 20 MAGs were only identified in AL1_15 and 25 MAGs were exclusively found in CH1_17. Clearly partitioned communities were observed between the two boreholes (Fig. 3 B). MAGs aligning with Asgardarchaeota, Bacteroidota, BMS3Abin14, Desulfobacterota, MBNT15, Myxococcota, Nitrospirota, Patescibacteria and Verrucomicrobiota were only identified in CH1_17, while members of phylum Firmicutes_A were exclusively detected in AL1_15.

The microbial communities also shifted with depth and age. In CH1_17, the number of recovered MAGs was highest in FCP_1.75 (n=22) and lowest in FCP_5.4 (n=4). The majority of the recovered populations from FCP were related to Chloroflexota_A at 1.75 m (40.53%, all quoted percentage values in the sections below refer to the relative abundance from the respective metagenome) and Actinobacteriota at 5.4 m (99.96%), whereas the recovered populations from MP at 16.9 m were mostly related to Chloroflexota and Gammaproteobacteria (34.93 and 29.21%, respectively). In addition, Acidobacteriota, Bacteroidota, Chloroflexota_A, Nitrospirota, Patescibacteria and Verrucomicrobiota MAGs were exclusively observed at
1.75 m, while BMS3Abin14, Desulfobacterota, MBNT15 and Myxococcota were only detected at 16.9 m. Lastly, archaeal MAGs aligning with Asgardarchaeota, which is suggested to be the closest prokaryotic relatives to eukaryotes [69, 70], were also only recovered from MP at 16.9 m. Another analysis of MAGs obtained from discrete depths of marine permafrost and annotated against TIGRFAM and COG [31] showed the presence of Asgardarchaeota, Bacteroidetes, Nitrospirae, and Deltaproteobacteria. Contemporary descendants of the microorganisms, which were exclusively discovered in CH1_17, were found to be highly metabolically flexible organisms that could adapt to resource variability by using different electron donors and acceptors or being involved in syntrophic interactions. For example, Myxococcota from anoxic aquatic environments being a strict anaerobe are capable of using fermentation, nitrate reduction, and dissimilarity sulfate reduction for energy acquisition [71]. The current discovery of MAGs for Desulfobacterota and MBNT15 in deeper MP_16.9 sample are in line with the previously obtained data that obligately anaerobic taxa such as sulfate reducers and candidate lineage MBNT15 thrive in more stable deeper marine sediments [72]. The 2 MAGs belonging to Asgardarchaeota were also discovered in the deeper marine sample. A recent review suggested that many Asgardarchaeota are involved in syntrophic interactions [73], e.g., the syntrophic exchange of formate and hydrogen was shown between a Lokiarchaeon and a sulfate-reducing Deltaproteobacterium [74].

In freshwater permafrost from AL1_15, the metagenome from 25.1 m generated the highest number of MAGs (n=19), followed by 6.0, 2.95 and 15.0 m (n=18, 8 and 4, respectively). Even though the FP_25.1 metagenome size was approximately half that of the FP_2.95 metagenome, it generated the highest number of MAGs. This could be explained by low biodiversity in sample FP_25.1 based on the 16S rDNA amplicon analysis (Fig. 1 A). While poor MAG recovery from younger permafrost samples could be related to high diversity of the microbial community [75]. A high abundance of the recovered populations was aligned with Chloroflexota_A at 2.95 and 6.0 m (73.73 and 44.27%, respectively), whereas MAGs recovered from 15.0 and 25.1 m were dominated by Actinobacteriota (59.16%) and Gammaproteobacteria (89.57%), respectively. In addition, populations affiliated with Chloroflexota_A and Gemmatimonadota were only identified at 2.95 and 6.0 m, while Firmicutes_A MAGs were exclusively detected at 25.1 m.

Amplicon and metagenome approaches detected similar populations of bacteria; however, amplicon analysis showed their presence in more samples than was detected by metagenomics. The number of MAGs correlated positively with microbial diversity based on amplicon analyses in FCP samples (higher number of MAGs in FCP_1.75 at higher Alpha diversity), and correlated negatively in FP samples (higher number of MAGs in FP_25.1 at lower Alpha diversity). This discrepancy may be connected to biases during isolation of DNA from samples with different salinity, sequencing approaches, depth of sequencing, platform for sequencing technology and bioinformatic approaches [76].

### 2.4 Metabolic Potential in Permafrost

...
Recovered MAGs were screened for genes encoding hydrolysis, fermentation, respiration, CO₂ and N₂ fixation, motility, bacterial secretion, spore formation and stress resistance (Fig. 4-5 and Tables S3-S5) to identify their metabolic potential. Genes coding for certain functions that were not identified in MAGs or in many samples were also searched in un-binned, assembled metagenomic sequences and a summary of potential metabolic pathways identified in permafrost metagenomes presented in Fig. 6.

2.4.1 Aerobic Respiration

Despite the scarcity of oxygen (<0.07 mg O₂ kg⁻¹ at redox potential below +200 mV [77]), genes encoding the machinery to reduce oxygen were identified in 30 MAGs, including members of Acidobacteriota, Actinobacteriota, Alphaproteobacteria, Bacteroidota, Chloroexota, Gammaproteobacteria, Gemmatimonadota, MBNT15 and Nitrospirota. Depending on the sample, 5.29-100% of the recovered populations contained genes coding for cytochrome c oxidase (Fig. 4).

These communities were most abundant at 15.0 and 25.1 m in FP (100 and 99.76%, respectively) and at 16.9 m in MP, respectively (35.69%). The high abundance of genes coding for cytochrome c oxidases could be explained by high identity of this enzyme in diverse organisms [78]. Cytochrome c oxidase catalyzes redox-driven proton pump that take part in generating the proton gradient in both prokaryotes and mitochondria that drives synthesis of ATP [78]. This highlighted the importance of energy acquisition through oxygen reduction in the permafrost populations, and accumulation of these enzymes in the deeper and older samples. In addition, MAGs containing genes encoding both low and high affinity cytochrome c oxidase had the highest relative abundance at 25.1 m in FP (92.49%) and at 16.9 m in MP (30.94%, Fig. 4), suggesting that the recovered populations from the deeper samples were capable of operating under different levels of O₂ concentration. This is in line with previous cultivation studies showing that isolates from the Siberian permafrost grew well at atmospheric oxygen concentrations [10, 79].

2.4.2 Carbohydrate Hydrolysis

Chitin and plant-derived materials, such as cellulose and xylan, are the two most abundant types of polysaccharides in the ecosystem [80]. MAGs containing genes encoding cellulases and/or β-glucosidases, which are involved in cellulose degradation, were most abundant in older permafrost at 15.0 and 25.1 m depth in FP (100 and 92.63%, respectively) and at 16.9 m in MP (80.85%) (Fig. 4) with carbon content of 1.896, 2.663 and 0.911%, respectively. These populations were dominated by Actinobacteriota in FP_15.0, Gammaproteobacteria in FP_25.1, and Chloroflexota and Gammaproteobacteria in MP_16.9 (Fig. 5).

MAGs with the capacity of carrying out xylan degradation were identified in all the samples apart from FP_2.95 and FCP_5.4. Approximately 10.99-31.31% of the recovered communities contained xylanase and/or β-xylosidase genes, with the highest at 15.0 m in FP (17.29%) and 16.9 m in MP (31.31%). These
potentially xylan degrading microorganisms were dominated by Actinobacteriota and Gammaproteobacteria, respectively. The higher abundance for cellulases over the protein-coding sequences for xylan breakdown shows that the Siberian permafrost populations have potential to degrade cellulose, the most commonly utilized polysaccharide. A recent study of Alaskan permafrost with ages of 19-33 kyr showed that sequences for enzymes targeting structural polysaccharides, such as xylan and cellulose, were less abundant than those targeting smaller molecular weight compounds [81]. However, the same study displayed that microbial community changed along the chronosequence from young permafrost community with potential to target hemicellulose, through increased potential to target starch, to the old permafrost microbial population having enzymes to target recalcitrant substrates like peptidoglycan and cellulose, that is in line with our findings. Populations encoding genes assigned to chitin degradation were most abundant at 2.95 and 15.0 m in FP (57.82 and 59.16%) and at 5.4 m in FCP (37.21%). MAG_99 and MAG_142, members of Actinobacteriota, were the dominating population for chitin degradation at 15.0 m in FP and at 5.4 m in FCP, respectively, whereas the dominant population at 2.95 m in FP belonged to one Chloroflexota_A MAG.

2.4.3 Fermentation

The potential for fermentation was identified in 23 MAGs, including members of an archaeal phylum, Asgardarchaeota. Microorganisms with the capability of producing acetate and/or lactate via fermentation were identified in all the samples. In FP sediments, fermentation-encoding microorganisms contributed a high abundance of the recovered communities at 15.0 and 25.1 m (100 and 91.76%, respectively), while approximately 5.29 and 12.02% of the recovered populations contained genes for fermentation at 2.95 and 6.0 m, respectively (Fig. 4). The most abundant MAGs encoding fermentation pathways were Gammaproteobacteria at 2.95, 6.0 and 25.1 m, whereas these populations primarily belonged to Actinobacteriota at 15.0 m (Fig. 5). Formate fermentation genes were only observed at 6.0 m in FP, whereas butanoate fermentation genes were exclusively identified at 25.1 m. The recent metagenomics study of the alluvial silty loams continuously frozen for 0.01-1.1 Myr showed that genes involved in the synthesis of formate, acetate, and butyrate were more numerically abundant in the older sediments [15]. In addition, MAGs containing genes indicating the capability for fermentative production of ethanol were identified at 6.0 and 25.1 m in FP and at 16.9 m in MP. In FCP and MP, genes encoding fermentation were most enriched in MP_16.9 (71.46%), followed by FCP_5.4 and FCP_1.75 (37.21 and 12.79%, respectively). Chloroflexota and Gammaproteobacteria MAGs appeared to be the most abundant groups with pyruvate metabolic genes at 16.9 m, while the most abundant fermentation encoding microorganisms were Bacteroidota at 1.75 m and Actinobacteriota at 5.4 m. Populations containing genes for formate and butanoate fermentation pathways were identified at 1.75 m in FCP and 16.9 m in MP. Analyses of neighboring samples showed presence of low–molecular-weight organic acids, such as formate, acetate, and propionate, at all depths with the highest concentrations of acetate (36.2 ± 3.3 μg g⁻¹) and formate (12.0 ± 0.7 μg g⁻¹) in FCP at depth of 5.8 m [31]. Lastly, one Myxococcota MAG containing genes encoding pyruvate fermentation to propanoate was exclusively identified in MP at 16.9 m. Overall, our data suggested that fermentation contributed significantly to anaerobic degradation of soil organic
carbon in the deeper and older permafrost (Fig. 4 and Table S4), which is in line with previous studies emphasizing its importance in permafrost [82, 83].

2.4.4 Potential Metabolic Interaction Between Populations

Short-chain fatty acids, such as alcohols, hydrogen and CO₂ converted by fermentative populations can subsequently be used by other microorganisms. Genes involved in ethanol degradation were detected in all the samples except for FCP_5.4. They were particularly abundant at 25.1 m in FP (82.63%) and at 16.9 m in MP (53.68%, Fig. 4). These populations were dominated by Gammaproteobacteria in FP_25.1 and MP_16.9 (Fig. 5). Hydrogen is an important electron donor for microorganisms in anaerobic conditions. MAGs containing homologs for anaerobic hydrogen oxidation were only identified in FP_25.1, FCP_1.75 and MP_16.9 (0.24, 7.43 and 3.15%, respectively). These populations belonged to Bacteroidota (n=1), Desulfobacterota (n=1) and Firmicutes_A (n=1). However, the NADP-dependent hydrogenase catalyzes reversible oxidation of H₂, therefore, the detected hydrogenase could also be involved in hydrogen production.

Five MAGs containing genes coding for form I ribulose-bisphosphate carboxylase (RuBisCO), the key enzyme in CO₂ fixation via the Calvin-Benson-Bassham (CBB) cycle and the most abundant form in eukaryotes and bacteria [84], were only identified in FP (Fig. S2 and Table S2). Of these populations, two Alphaproteobacteria MAGs and two Gammaproteobacteria MAGs contained both \( cbbL \) and \( cbbS \) genes for form I RuBisCO, while the Chloroflexota_A MAG only included the \( cbbL \) gene. These microorganisms were most abundant in FP_6.0 (22.84%) and were mainly encoded by Chloroflexota_A (Fig. 4 and 5). In addition, three archaeal MAGs, from Asgardarchaeota and Crenarchaeota, were suggested to hold \( cbbL \) genes encoding form III RuBisCO and were only identified in MP at 16.9 m (10.59%). This form is known to be present in many archaea (Jaffe et al., 2019) and reported to be involved in the incorporation of CO₂ into ribulose-1,5-bisphosphate (RuBP) from nucleotides like adenosine monophosphate (AMP) [85, 86]. Genes coding for form IV RuBisCO were identified in 3 MAGs, including the two MAGs also containing genes for form I and one archaeal MAG also with genes for form III. Notably, form IV, which is often referred to as a RuBisCO-like protein, appears to be involved in sulfur metabolism, methionine salvage pathway, and D-apiose catabolism [87, 88], rather than the CBB cycle. The microorganisms carrying this form were found at 2.95, 6.0 and 25.1 m in FP and 16.9 m in MP. MAG_40, closely related to Nitrospirota, encoded genes assigned to CO₂ fixation via incomplete reductive tricarboxylic acid (TCA) cycle, and was exclusively identified in FCP at 1.75 m (3.38%). Lastly, key genes encoding both anaerobic carbon-monoxide dehydrogenase and acetyl-CoA synthase for the Wood-Ljungdahl pathway were found in 4 MAGs. The Actinobacteriota MAG was only identified in FCP at 1.75 m (4.19%), whereas the Desulfobacterota and Chloroflexota MAGs were detected in MP at 16.9 m (33.64%). Overall, the higher abundance of CO₂ fixation potential at 16.9 m in MP (33.64%) suggested that autotrophy was more widespread when this permafrost layer formed relative to the others.
2.4.5 Methane Metabolism

Methane is a potential electron donor for anaerobic respiration, coupled to sulfate, iron, manganese, nitrate, nitrite reduction or denitrification [89-95]. It is also a potent greenhouse gas that contributes to global warming. We screened genes encoding methanogenesis, anaerobic and aerobic methane oxidation from the recovered MAGs as well as un-binned metagenomic sequences. None of the MAGs were suggested to carry genes involved in anaerobic methane oxidation (AOM) or methanogenesis. Within the un-binned metagenome data, the key genes involved in methanogenesis (based upon the presence of the *mcrBCDG* genes) were identified in all samples except FP_15.0 and FCP_1.75 (Fig. S3). These genes were most enriched in FCP_5.4 (367.46 TPM), corresponding to the 16S rRNA amplicon data showing the highest relative abundance of methanogenic taxa, and correlated with a presence of methane in layer between 4.8 and 6.4 m with the highest concentration of methane 76 mmol kg\(^{-1}\) detected near this depth at 5.6 m. The concurrent presence of genes involved in methanogenesis, ASVs associated with methanogens and methane detected by the static headspace method [67] suggests that methane at 5.6 m depth has a biological origin and methane production may happen in situ at FCP conditions. Overall, the capacity for methanogenesis appeared to be relatively low in the remaining samples (0.006-7.81 TPM) which is in line with the previous studies showing that methanogenesis was extremely limited in intact permafrost [17, 96]. Despite high levels of methane 157.1 and 167.2 mmol kg\(^{-1}\) detected in FP at 15.0 and 25.1 m, respectively [67], the key genes for methanogenesis were not identified at 15.0 m and were negligible at 25.1 m (0.006 TPM). A previous study, which sampled permafrost from the same location at three depths (1.4, 11.8, and 24.8 m), also showed low abundance of methanogens [15]. The low abundance of methanogenesis genes in samples with high concentrations of methane that has biogenic origin (-84.9 ± 0.5‰ for \(d^{13}C_{CH_4}\)) may suggest that this methane has either surficial or deep-sediment origin and likely accumulated in lithological traps earlier during permafrost formation [67] or represents a consequence of methanogenesis in deeper permafrost layers, similar to observations from organic-rich Antarctic marine sediments where the methanogenesis genes are in low abundance despite the observation of biogenic methane [97]. Low abundance of methanogens in metagenomes from permafrost samples with biogenic methane may also be attributed to biases associated with sample size, DNA extraction, sequencing technology, depth of sequencing, sequence assembly, annotation, and database used for identification [76]. Analysis of replicate metagenomes and ultra-deep sequencing of additional samples from methane containing permafrost layers would likely result in more prominent detection of methanogens.

In contrast to the rare occurrence of anaerobic methane oxidation in permafrost, aerobic methane oxidation typically attenuates methane release [98, 99]. However, MAG_40, encoding the capacity to oxidize methane aerobically, was only found in FCP at 1.75 m (most closely to Nitrospirota, 3.38%, Fig. 4). Within the un-binned metagenomic sequences, genes involved in aerobic methane oxidation (based on the presence of the *pmoABC-amoABC*) were identified at 2.95 and 6.0 m in FP (1.38 and 4.06 TPM, respectively, Fig. S3) and at all depths in FCP and MP, with highest abundance in FCP_1.75 (15.05 TPM). The generally low abundance of genes for aerobic methane oxidation was not surprising considering the
lack of oxygen in permafrost. In agreement with the high abundance of methanogenesis genes at 5.4 m in FCP, genes for aerobic methane oxidation were lowest at this depth (0.77 TPM), indicating the strictly anaerobic microenvironments at this depth. It should be noted that the identified genes encoding aerobic methane oxidation were also suggested to co-oxidize ammonia aerobically [100].

2.4.6 Nitrogen and Sulfur Metabolism

The potential ability to fix nitrogen was not identified in the recovered MAGs. Within the un-binned metagenome data, nitrogenase genes were identified in all samples with the exception of FP_15.0. An overall low abundance of form I and III \( \text{nifH, nifD, and nifK} \) genes was observed across the samples (totaling 0.12-16.95 TPM, Fig. S3 and S4). Genes for the enzymes needed to utilize nitrate or nitrite as a terminal electron acceptor in the process of nitrification were detected in 26 MAGs. Populations that were suggested to reduce nitrate to either nitrite or ammonium as the end products were most abundant at 25.1 m in FP (62.84%) and at 16.9 m in MP (48.46%) that correspond to higher concentration of ammonium in deeper sediments (~300 ppm \( \text{NH}_4^+ \) g\(^{-1}\) at depth of 12-24 m in FP and 110-152 ppm \( \text{NH}_4^+ \) g\(^{-1}\) in MP [59]). These populations were dominated by Gammaproteobacteria in both samples. Genes encoding the capacity to reduce nitrite to ammonium were also most enriched at 25.1 m in FP (34.13%) and were identified at 1.75 m in FCP with a relatively lower abundance (5.45%). In addition to the dissimilatory nitrate reduction pathway, three MAGs also contained a complete or near complete set of genes coding for denitrification. These populations represented a small fraction of the recovered communities, which is present in all the samples (0.03-4.74%), except for FP_15.0 and FP_25.1 (40.84 and 58.19%, respectively). The presence of nitrite and nitrate in different FP layers and in deeper MP layers supports a potential of the communities to the denitrification process. No MAGs were suggested to carry out denitrification in FCP_5.4 and nitrate was not detected in FCP layer. Overall, our data suggest that nitrogen metabolism plays an important role in the older permafrost, likely due to nitrate being the most energetically favorable electron acceptor in the absence of oxygen. The generally higher abundance of populations involved in nitrate reduction is consistent with previous reports of their dominance under reducing, high-carbon and low-nitrate conditions [101, 102].

Dissimilatory sulfate reduction by anaerobic microorganisms is a predominant pathway of organic material mineralization in marine sediments [103]. However, none of the MAGs contained the key \( \text{dsrA and/or dsrB} \) genes for dissimilatory sulfate reduction, suggesting the rare utilization of this pathway by permafrost populations in the studied environments. Populations containing genes for enzymes needed to reduce sulfur via polysulfide were exclusively identified in MP at 16.9 m (51.08%). The presence of hydrogen sulfide in MP samples was identified by the specific odor during core extraction in the field. These populations were encoded by Asgardarchaeota, Crenarchaeota and Chloroflexota. Genes encoding thiosulfate oxidation via the Sox pathway were only identified in FP at 15.0 and 25.1 m (0.34 and 6.35%, respectively). Populations containing these genes exclusively belonged to Alpha- and Gammaproteobacteria, where the Alphaproteobacteria MAGs were only identified in FP_25.1. In agreement with the genomic analysis, an overall low abundance of the \( \text{dsrAB} \) genes coding for sulfate
reduction were detected at 2.95 and 6.0 m in FP, 1.75 and 5.4 m in FCP and 16.9 m in MP (1.57-18.68 TPM) within the un-binned metagenome data. Nevertheless, the possibility of the sulfate reduction was supported by the presence of sulfate in FP (4.4-8.8 mmol kg\(^{-1}\), except sample FP_25.1 where sulfate was ~10 times lower), FCP (1.7-2.7 mmol kg\(^{-1}\)), and MP (8.1-36.8 mmol kg\(^{-1}\)) samples. The content of sulfate in all studied samples was <0.2% by mass what is considered low sulfate soils [104]. The previous study also discovered the presence of sulfate reduction in marine and deep freshwater permafrost sediments with S\(^2-\) in concentration of 0.12-0.22 and 0.12-0.35 g kg\(^{-1}\) wet soil, respectively [27]. The phylum Desulfobacterota that comprises sulfate-, sulfur-, and ferric iron-reducing bacteria was solely detected in MP at 16.9 m. In addition, genes encoding the use of polysulfide sulfur were more abundant in MP within the un-binned metagenome data (Fig. S3), suggesting that anaerobic degradation of organic carbon coupled to sulfur reduction was more common in MP. Of the aforementioned genomes that encoded the capacity for fermentation and/or anaerobic respiration, a high abundance of the recovered communities also contained genes encoding the machinery to reduce oxygen at 15.0 and 25.1 m in FP (100 and 96.96%, respectively). Taking these together, we suggest that the recovered communities from the older samples of FP are likely to be facultatively anaerobes, whereas the recovered populations from 5.4 m in FCP to 16.9 m in MP are more likely to be strict anaerobes at the time of freezing (Table S4).

### 2.4.7 Potential Adaptation in Permafrost

Biofilm formation is considered to be a survival strategy to enable adaptation of microorganisms to extreme environments [105], i.e., bacteria were tightly associated with soil particles in Siberian permafrost [106]. In addition, *Psychrobacter arcticus* strain 273-4 from Siberian permafrost showed the capability to form biofilms [107]. Genes coding for both surface attachment through flagellar (including chemotaxis) and/or type IV pili and extracellular polymeric substance (EPS) secretion via type II secretion system (T2SS) which are potentially involved in biofilm formation were identified across seven MAGs representing Gammaproteobacteria, MBNT15 and Myxococcota. Biofilm-forming Gammaproteobacteria were most abundant at 25.1 m in FP (73.10%) and decreased in abundance toward the surface (Fig. 4). In FCP and MP, the highest abundance of the biofilm-forming microorganisms was detected in MP_16.9 (33.07%). None of the MAGs from FCP_5.4 contained genes encoding both surface attachment and EPS secretion. Of the 7 MAGs, one Gammaproteobacteria MAG also included genes for type IV secretion system (T4SS) which encodes conjugation machinery and DNA release and uptake systems [108]. Another three Gammaproteobacteria MAGs, MAG_24, MAG_35 and MAG_67 also contained genes coding for type VI secretion system (T6SS, delivering toxins into eukaryotic and prokaryotic cells, [109]), whereas genes encoding type III secretion system (T3SS, injecting effector proteins into eukaryotic cells, [110]) were also detected in MAG_35. In addition, three MAGs included genes for flagellar and/or type IV pili assisted motility and T6SS, where one MAG also contained genes encoding T4SS. These microorganisms were exclusively identified in FP at 25.1 m (3.80%). MAGs containing genes for flagellar or type IV pili mediated motility but lacking genes encoding secretion systems were also identified at 6.0 and 25.1 m in FP (0.12 and 4.89%, respectively), 1.75 m in FCP (8.81%) and 16.9 m in MP (5.12%). In addition, approximately 0.02-14.24% of the recovered MAGs contained genes assigned to type I secretion
system (T1SS), T2SS or T6SS at 2.95, 6.0 and 25.1 m in FP and 1.75 m in FCP, with the highest abundance observed in FP_25.1. The high abundance of recovered MAGs that contained genes encoding both surface attachment and bacterial secretion system in FP_25.1 and MP_16.9 indicated that biofilm formation might be an important survival strategy for microbes in the older and deeper perennially frozen sediments, providing exchange of molecules and ions between live microorganisms and liquid brine veins surrounding cells and soil particles [5, 16]. This is generally in agreement with the un-binned metagenome data analysis where genes involved in chemotaxis, flagellar assembly, type IV pili and bacterial secretion systems (type I-III and VI) were more abundant in the oldest permafrost sample (FP_25.1), while genes encoding T4SS were most abundant in MP_16.9 (Fig. S5). The higher abundance of T4SS in marine permafrost could indicate that horizontal gene transfer may play a role in allowing microorganisms to adapt to changes in their environment. Overall, our data support the findings from a previous metagenome study of Alaska permafrost showing that chemotaxis and bacterial secretion system pathways were enriched in older permafrost up to 33 kyr [60], but also pointed to the presence of bacterial populations that were potentially involved in biofilm formation.

Sporulation is a widely utilized strategy for microorganisms to survive in extreme environmental conditions [111]. MAG_104, most closely related to Firmicutes_A, was suggested to carry out multiple stages of spore formation and was exclusively found in FP at 25.1 m (0.24%, Fig. 4 and 5). The low relative abundance of spore-forming populations in the older permafrost contradicted a previous study from the same location showing that the older layers in freshwater permafrost was dominated by spore-forming bacteria [15], but was consistent with other studies of Siberian and Antarctica permafrost [112, 113] as well as permafrost from Svalbard, Norway [114]. Our un-binned metagenome data analysis also showed that metagenomes from 15.0 and 25.1 m had a lower abundance of these genes compared to the top layer samples in FP (Fig. S5), suggesting that spores were not the most prominent survival strategy in the older permafrost. The discrepancy between this study and the previous study [15] of freshwater permafrost suggested that the permafrost sampled at this site may be highly heterogeneous, providing large amounts of micro niches with different environmental characteristics, even from the same borehole.

Since permafrost is frozen, genes encoding cold shock protein (CSPs) were prevalent across the recovered MAGs (38 of 60 MAGs). CSPs are a loosely defined group of DNA binding proteins that are commonly found in cold-adapted microorganisms and were most enriched at 15.0 m in FP (100%), followed by 25.1, 6.0 and 2.95 m (99.76, 67.74 and 62.95%, respectively, Fig. 4). In FCP and MP, CSP genes were highest in FCP_5.4 (100%, respectively) and lowest in FCP_1.75 m (25.46%). The overall high abundance of CSP genes across the samples indicates that microbial communities are well-adapted to stresses associated with freezing temperatures, nutrient starvation and growth deprivation [115].

**Conclusion**

This study showed that age, depth and salinity of permafrost sediments shaped biodiversity, community composition and metabolic potential in ancient Siberian permafrost deposits. Microbial communities
shifted from Actinobacteriota at the top layer to Gammaproteobacteria and Chloroflexota at the deeper layers in AL1_15 and CH1_17, respectively. In addition, genome- and gene-centric analyses revealed that the recovered populations encoded the capacity for hydrolysis, fermentation, dissimilatory nitrate reduction, denitrification, and sulfur reduction (Fig. 6).

Despite many differences between the freshwater coastal permafrost (FCP) and freshwater permafrost (FP), their microbial communities were similar compared with deep old marine permafrost (MP) and deep old freshwater permafrost (FP). However, many functions, especially fermentation and nitrogen cycling, could be performed by many members of each of these communities, and were present in all samples. Environmental characteristics, especially temperature, affect nearly every aspect of how microorganisms interact with and are constrained by their environment. An average microbial metabolic rate estimated at permafrost temperatures of -5° to -10°C is low and equals $10^{-2}$-$10^{-4}$ g C h$^{-1}$ [116]. The metabolic processes are susceptible to changes in temperature and metabolic rate, being temperature-dependent, is expected to increase with increasing environmental temperature. With ongoing climate warming a threat of abrupt thawing of glaciers, ground ice, and permafrost increases. In its turn the permafrost thawing will cause a revival of permafrost microbes and trigger decomposition of millennia old carbon stocks. Analyses of permafrost metagenomes unveil taxonomic and metabolic potential of preserved at negative temperatures microbial communities and help to predict their behavior upon permafrost thawing. Thus, the current study showing the widespread presence of cytochrome c oxidase in Siberian permafrost suggested that microbial communities with potential for aerobic respiration will function well in oxygenated soils. Similar to our findings, a recent study of Svalbard permafrost metagenomes found that MAGs were enriched in genes involved in aerobic respiration and regulation of ammonium, and sulfur [114]. Permafrost microorganisms exhibited diverse metabolic potential and could be shaped into either aerobic or anaerobic communities under influence of physicochemical properties and environmental factors upon permafrost thawing.

**Declarations**

**Data Availability Statement**

The datasets generated for this study can be found in the NCBI Short Read Archive (SRA) under accession number PRJNA634390 for amplicon sequencing data and under accession number PRJNA601698 for metagenome raw reads.

**Originality-Significance Statement**

Knowledge about microbial diversity and metabolism in ancient permafrost is very limited. We studied changes in biodiversity and functional gene distributions with response to age, depth and salinity of Siberian permafrost. We concluded that although there are large differences in microbial community compositions between young vs. old and freshwater vs. marine deposits, the functional possibilities for these communities are similar, especially with respect to carbon degradation.
Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

X.W., A.A. and D.W. performed DNA extraction, library preparation, sequencing, and initial quality filtering. X.W. performed computational work, and wrote the manuscript. S.P., E.R., K.L., and T.V. were responsible for funding acquisition, contributed to the interpretation of the data and editing of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the National Science Foundation DEB-1442262 (to TAV, KGL), the Russian Foundation for Basic Research (RFBR) 19-29-05003mk (to EMR), and partially supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under award number DE-SC0020369 (to TAV, KGL).

Acknowledgments

We thank Andrey A. Abramov and Victor A. Sorokovikov for overseeing sample collection in the northeastern Siberia; and Katie M. Sipes, Wyatt A. Cyr and Oksana G. Zanina for help with physicochemical analyses. We would like to acknowledge a substantial contribution of Tullis C. Onstott, who passed away on October 19, 2021.

References

Astrobiology 2003, 3(2):331-341.


**Figures**

**Figure 1**

Alpha diversity measured by Shannon and Simpson indexes (A) and NMDS ordination plot (B) based on 16S rRNA gene amplicon sequencing.
Figure 2

Microbial community compositions at the phylum level (class level for Proteobacteria) determined by 16S rRNA amplicon sequencing. Phyla with abundance <1% were grouped into ‘Other’.
Figure 3

Distribution of the recovered MAGs and their relative abundance. (A) Heatmap showing relative abundance of the recovered MAGs. The bar plots show the total percentage of sequencing reads mapped to the MAGs and the total number of MAGs recovered in the corresponding sample. Numbers given in brackets indicate the number of MAGs assigned to the corresponding taxon. The zero written in the heatmap indicates the absence of that MAG. (B) Network showing the number of MAGs recovered from each sample of the two sampling sites. Color codes for the suggested taxonomy are shown as in A.
Figure 4

Heatmap showing the total relative abundance of MAGs containing genes encoding the corresponding metabolic function from the respective samples.
Figure 5

Genome-enabled metabolic potential per taxonomic distribution of freshwater (A) and freshwater coastal and marine permafrost (B). The plots show the total relative abundance of different function and taxon from the respective metagenomes. Only the sample with highest relative abundance is shown for each function and taxon. Full taxonomic ranges of all described functions are presented in Table S3.
Figure 6

Flow diagram of the potential metabolic pathways derived from gene assignments from the MAGs reconstructed in this study and the un-binned metagenome data. Arrows represent metabolic capabilities that were identified in this study. The dashed lines represent potential electron donors for the anaerobic respiration processes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- WuSupplemTables.xlsx
- WuSupplementaryfigures.docx