microLife, 2022, **3**, 1–15
DOI: 10.1093/femsml/uqac003
Advance access publication date: 7 April 2022
Research Article

Metagenomic survey of the microbiome of ancient Siberian permafrost and modern Kamchatkan cryosols

Sofia Rigou[1], Eugène Christo-Foroux[1], Sébastien Santini[1], Artemiy Goncharov[2], Jens Strauss[3], Guido Grosse[3,4], Alexander N. Fedorov[5,6], Karine Labadie[7], Chantal Abergel[1], Jean-Michel Claverie[1,*]

[1]IGS, Information Génomique & Structurale (UMR7256), Institut de Microbiologie de la Méditerranée (FR 3489), CNRS, Aix Marseille University, Marseille, 13288, France
[2]Department of Molecular Microbiology, Institute of Experimental Medicine, Saint Petersburg, Russia, Department of Epidemiology, Parasitology and Disinfectology, Northwestern State Medical Mechnikov University, Saint Petersburg, 195067, Russia
[3]Permafrost Research Section, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 14473 Potsdam, Germany
[4]Institute of Geosciences, University of Potsdam, 14478 Potsdam, Germany
[5]Melnikov Permafrost Institute, 677010 Yakutsk, Russia
[6]BEST International Centre, North-Eastern Federal University, 677027 Yakutsk, Russia
[7]Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, Évry, 91000, France
*__Corresponding author:__ Laboratoire Information Génomique et Structurale (IGS) UMR7256, Aix Marseille Université, CNRS, Parc Scientifique de Luminy – 163 Avenue de Luminy, 13288, Marseille cedex 09, France. Tél: 04 13 94 67 77; E-mail: Jean-Michel.Claverie@univ-amu.fr
One sentence summary: Permafrost bacteria appear to constitute an enormous reservoir of antibiotic resistance genes.
Editor: Julian Parkhill

Abstract

In the context of global warming, the melting of Arctic permafrost raises the threat of a reemergence of microorganisms some of which were shown to remain viable in ancient frozen soils for up to half a million years. In order to evaluate this risk, it is of interest to acquire a better knowledge of the composition of the microbial communities found in this understudied environment. Here, we present a metagenomic analysis of 12 soil samples from Russian Arctic and subarctic pristine areas: Chukotka, Yakutia and Kamchatka, including nine permafrost samples collected at various depths. These large datasets (9.2×10^{11} total bp) were assembled (525 313 contigs > 5 kb), their encoded protein contents predicted, and then used to perform taxonomical assignments of bacterial, archaeal and eukaryotic organisms, as well as DNA viruses. The various samples exhibited variable DNA contents and highly diverse taxonomic profiles showing no obvious relationship with their locations, depths or deposit ages. Bacteria represented the largely dominant DNA fraction (95%) in all samples, followed by archaea (3.2%), surprisingly little eukaryotes (0.5%), and viruses (0.4%). Although no common taxonomic pattern was identified, the samples shared unexpected high frequencies of β-lactamase genes, almost 0.9 copy/bacterial genome. In addition to known environmental threats, the particularly intense warming of the Arctic might thus enhance the spread of bacterial antibiotic resistances, today's major challenge in public health. β-Lactamases were also observed at high frequency in other types of soils, suggesting their general role in the regulation of bacterial populations.

Keywords: metagenomics, Kamchatka, Yakutia, Siberia, permafrost thaw, antibiotic resistance genes

Introduction

Two of the front-page societal concerns are global warming and the increasing frequency of emerging infectious diseases, eventually turning into pandemics (Morens *et al.* 2020, Zhou *et al.* 2020). It turns out that both predicaments are partly linked. Through its influence on the ecology of living organisms vectoring infectious pathogens, the warming climate contributes to spreading endemic diseases from tropical regions into temperate areas of the globe, such as Western Europe (Hertig 2019, Fuehrer *et al.* 2020, Liu *et al.* 2020). This phenomenon is further amplified by the encroachment into pristine areas (e.g. tropical forests) where expanding human activities generate new risks through imponderable contacts with (mostly uncharted) microbial environments and their associated wildlife hosts (Bradley and Altizer 2007, Keita *et al.* 2014, Plowright *et al.* 2017, Valentine *et al.* 2019).

While such dangers are mostly pointed out as coming from the South, more recent concerns have been raised that new plagues could also come from the Arctic, through the release of infectious agents until now trapped in perennial frozen soils (i.e. permafrost) up to 1.5 km deep and 2–3 million years old (Revich and Podolnaya 2011, Revich *et al.* 2012, Parkinson *et al.* 2014, Huber *et al.* 2020).

Climate warming is particularly noticeable in the Arctic where average temperatures are increasing twice as fast as in temperate regions (Cohen *et al.* 2014). One of the most visible consequences is the widespread thawing of permafrost at increasing depths (Biskaborn *et al.* 2019, Turetsky *et al.* 2019) and the rapid erosion of permafrost bluffs (Fuchs *et al.* 2020), a phenomenon most visible in Siberia where deep continuous permafrost underlays most of the North Eastern territories.

The 'ice age bug' concern has been periodically brought back to the public attention. For instance, when an exceptionally hot summer triggered local outbreaks of anthrax on Yamal Peninsula, Northwest Siberia, in 2016, a deeper than usual summer season thaw of soils above the permafrost layer (i.e. the 'active layer') exhumed infectious *B. anthracis* endospores buried in the frozen ground for 75 years (Timofeev *et al.* 2019). Historically frequent

Received: November 30, 2021. **Revised:** March 11, 2022. **Accepted:** March 25, 2022

Table 1A. Deep permafrost samples analyzed in this study.

Sample	IGS code	Latitude (°)	Longitude (°)	Locality	Depth (m)	Dating years BP	Type
P 1084T	B	68.37016	161.41555	Stanchikovskiy Yar	NA	35 000	Wall sampling, Yedoma bluff
Yuk15 Yed1/Yed.6	K	61.75967	130.47438	Yukechi Alas landscape	16	>49 000	Core, Yedoma upland
Yuk15 Yed1/Yed.7	P				11	>36 000	
Yuk15 Alas1/Ala.11	L	61.76490	130.46503		12	>28 000	Core, from a drained thermokarst lake basin (Alas)
Yuk15 Alas1/Ala.10	N				16	>45 000	
Yuk15 Alas1/Ala.9	R				19	>42 000	
Yuk15 YuL15/Y1	O	61.76086	130.47466		19	>40 000	Core, Yedoma upland, under lake (4 m water depth)
Yuk15 YuL15/Y2	M				16	>48 500	
Yuk15 YuL15/Y4	Q				6	53[a]	

[a]: This soil sample was not frozen (talik). O, M, Q sample data from: Jongejans LL, Grosse G, Ulrich M *et al.* (2019). Radiocarbon ages of talik sediments of an alas lake and a Yedoma lake in the Yukechi Alas, Siberia. PANGAEA, https://doi.org/10.1594/PANGAEA.904738.

outbreaks of anthrax killed 1.5 million reindeer in Russian North between 1897 and 1925 and human cases of the disease have occurred in thousands of settlements across the Russian North (Hueffer *et al.* 2020).

The capacity of deeper (hence much older) permafrost layers to preserve the integrity of 'live' (i.e. infectious) viral particles was demonstrated by the isolation of two previously unknown giant DNA viruses: Pithovirus (Legendre *et al.* 2014) and Mollivirus (Legendre *et al.* 2015). Those Acanthamoeba-infecting viruses were isolated (and cultivated) from a wall-sampled layer of a well-studied permafrost bluff (the Stanchikovsky Yar) (sample B, Table 1A) radiocarbon dated to 35 000 years before present (BP) (Legendre *et al.* 2014). This first demonstration that eukaryotic DNA viruses could still be infectious after staying dormant since the late Pleistocene came after many other studies (Vishnivetskaya *et al.* 2006, Hinsa-Leasure *et al.* 2010, Graham *et al.* 2012) going back to 1911 (reviewed by Gilichinsky and Wagener 1995). It was shown that ancient bacteria could be revived from even much older permafrost layers, up to half a million years (Johnson *et al.* 2007). Contamination by modern bacteria has been suspected in the case of older isolates (Willerslev *et al.* 2004). A flower plant was also revived from 30 000-year-old frozen tissues (Yashina *et al.* 2012), as well as various fungi (from Antarctica) (Kochkina *et al.* 2012) and amoebal protozoans (Shmakova *et al.* 2016, Malavin and Shmakova 2020). There is now little doubt that permafrost is home to a large diversity of ancient microorganisms that can potentially be revived upon thawing. Many of them, in particular among viruses, are unknown. If detected, the follow-up questions are: Are those microorganisms a threat for today's society? Could some of them represent an infectious hazard for humans, animals or plants? Can we identify microbes the metabolisms of which would accelerate the emission of greenhouse gases (e.g. methane and carbon dioxide) upon thawing?

Beyond the direct infectious risk represented by viable microorganisms released from thawing ancient permafrost layers, the persistence of a multitude of genomic DNA fragments from all the organisms that were present in the soil at various times is also to be considered. Although it originated from dead (or even extinct) organisms, the permafrost DNA content constitutes a historical library of genetic resources, from which useful elements (genes) can be reintroduced into modern organisms through bacterial transformation and/or lateral gene transfer in protozoa or viruses. Of particular interest are DNA fragments coding for bacterial virulence factors or antibiotic resistance genes (ARGs) the origin of which predates by far human history (Finley *et al.* 2013, Perron *et al.* 2015, Potron *et al.* 2015, Wright 2019).

To address these questions, a first step is to evaluate the DNA in frozen soils. The technique of choice (which avoids the risk of reactivating unknown dangerous pathogens) for this task is metagenomics. In contrast with targeted approaches, such as metabarcoding (i.e. PCR-amplified subset of environmental DNA), it offers the possibility of unexpected discoveries in all domains of life and provides an estimate of the relative abundances of the various organisms, although at the cost of a lower sensitivity. Finally, the metagenomic analysis of DNA obtained from up to 60 g of frozen soil samples (Table S1, Supporting Information) is much less sensitive to eventual contaminations by minuscule amounts of ambient DNA or contemporary microorganisms and more suitable for site-to-site comparisons than culture-based approaches.

Here, we report the analysis of nine subsurface Siberian permafrost samples, and of three modern surface soils from pristine area in Kamchatka. All the permafrost samples were taken from carbon-rich, ice-rich frozen soils called Yedoma deposits (Schirrmeister *et al.* 2013, Strauss *et al.* 2017) found in vast regions of northeast Siberia (Lena and Kolyma river basins in Yakutia) (Fig. S1, Supporting Information) and known for well-preserved Late Pleistocene megafauna remains (mammoths, wholly rhinoceros). Yedoma permafrost is of special interest as it is prone to rapid thaw processes such as thermokarst and thermo-erosion (Nitzbon *et al.* 2020) releasing not only carbon as greenhouse gases (Schneider von Deimling *et al.* 2015, Turetsky *et al.* 2020) but also its formerly freeze-locked microbial content, in which we detected a high abundance of β-lactamases genes carried in the genomes (or plasmids) of a large phyletic diversity of bacteria.

Methods

Sample collection and preparation

We sampled Yedoma and thermokarst sediments from Central Yakutia from the Yukechi Alas landscape ~50 km southeast of Yakutsk (Table 1A). The Yukechi landscape is characterized by Yedoma uplands and drained as well as extant lake basins, so-called Alases, indicating active permafrost degradation processes (Fedorov and Konstantinov 2003). Field work took place in March

2015 during a joint Russian-German drilling expedition. We used a Russian URB2-47 drilling rig with a steel core barrel without a core catching system. Core barrels had different diameters, starting with 15.7 cm for the upper core segments and narrowing down to 8 cm for lower core segments, a setup preventing to get stuck in the deep borehole. To avoid contaminations, coring was conducted as dry drilling and no drilling fluid was used. The retrieved core segments were pushed out from the core barrel with air pressure only. The core segments were visually described and immediately stored in plastic bags or core foil and sealed tightly with duct tape. The wrapped and sealed cores were put into opaque and thermally insulated hard case expanded polypropylene boxes (thermoboxes) and stored outside, where they were kept frozen (or refroze slowly if the sediment material was originally unfrozen) at outside temperature around −10°C to −20°C. During the entire transportation, the core material was kept in the dark and frozen in the thermoboxes. Temperatures below freezing during the entire storage and transport chain were recorded with a temperature data logger in one of these boxes from the drilling site to the laboratories in Potsdam, Germany.

In total, three long (all ∼20m below surface/lake ice) permafrost sediment and unfrozen thaw bulb (talik) cores were retrieved: one from Yedoma deposits (61.75967°N, 130.47438°E), one below a lake on the Yedoma upland (61.76086°N, 130.47466°E) and one from the adjacent drained Yukechi alas basin (61.76490°N, 130.46503°E) (Table 1A; Fig. S1, Supporting Information). Biogeochemical characterization (Windirsch *et al.* 2020, Ulrich *et al.* 2021) as well as greenhouse gas production (Jongejans *et al.* 2021) is published. In Potsdam, the frozen cores were handled in a freezer laboratory at −10°C. They were split lengthwise using a band saw and were subsequently subsampled. After cleaning the splitting surfaces by scratching with cleaned knifes, the core stratigraphies were described visually and photographed. The sampling core halves were cut into subsamples and divided for further analysis, while the archive core halves were wrapped and sealed again for potential future additional analysis. For the samples used in this study, we took the inner part of the cores only to avoid any other sources of contamination on the outer core margins. After extraction, the subsamples were sent frozen with dry ice to Marseille, France, for further analyses.

The Stanchikovsky Yar sample (sample B) from which two live viruses were previously isolated (Legendre *et al.* 2014, 2015), and three modern surface soils from pristine cold regions (Kamchatka) were also analyzed for comparison (Table 1B). Figure S1 (Supporting Information) indicates the locations of these sampling sites.

DNA extraction

Samples were processed in BSL2 conditions using either the Miniprep (0.25 g of starting material) or Maxiprep (20 g of starting material) version of the DNeasy PowerSoil Kit from Qiagen (Venlo, Netherlands). Extractions were repeated until ∼1 μg (or more) of raw DNA was obtained for each sample (Table S1, Supporting Information). Samples B, C, D, E were processed using the miniprep kit, following the manufacturer's protocol and by adding dithiothréitol (DTT) to the C1 buffer at a 10 mM final concentration. All other samples (K, L, M, N, O, P, Q, R) were processed using the Maxiprep kit, including the addition of DTT to the C1 buffer at a 10 mM final. Samples were ground for 20 s in an MP FastPrep homogenizer (MP Biomedicals, Santa Ana, California, USA) at a speed of 4 m/s then incubated for 30 min at 65°C and finally ground again for 20 s at 4 m/s. After elution in 5 ml of elution buffer the extracted raw DNA was concentrated on a silica column from the Monarch Genomic DNA purification Kit from New England Biolabs (Ipswich, MA, USA). The DNA contents were then quantified using Qubit (Thermo Fisher Scientific, Waltham, MA, USA) and NanoDrop (Thermo Fisher Scientific) assays (Table S1, Supporting Information) prior sequencing.

DNA sequencing, assembly and annotation

Each metagenomic sample was sequenced with DNA-seq paired-end protocol on Illumina (San Diego, CA, USA) HiSeq 4000 platform at the French National Sequencing Center 'Genoscope' (https://jacob.cea.fr) producing datasets of 2 × 150 bp read length, except for sample B (2 × 100 bp read length) (Table S1, Supporting Information). Sequencing libraries were prepared as previously described in Alberti *et al.* (2017) using an 'on beads' protocol. Briefly, when available, 250 ng of genomic DNA were sonicated using the E210 Covaris instrument (Woburn, Massachusetts, USA) and the NEBNext DNA Modules Products (New England Biolabs) were used for end-repair, 3′-adenylation and ligation of NextFlex DNA barcodes (Bioo Scientific Corporation, Austin, TX, USA). After two consecutive 1× AMPure XP cleanups (Beckman Coulter, Brea, CA, USA), the ligated product were amplified by 12 cycles of PCR using Kapa Hifi Hotstart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA), followed by 0.6× AMPure XP purification. Finally, libraries were quantified by quantitative PCR (qPCR) (MxPro; Agilent Technologies, Santa Clara, CA, USA) and library profiles were evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies).

Reads were quality checked with FASTQC (v0.11.7) (Andrews 2014). Identified contaminants were removed and remaining reads were trimmed on the 3′ end using 30 as quality threshold with BBTools (v38.07) (BBMap 2022). Assemblies of filtered datasets were performed using MEGAHIT (v1.1.3) (Li *et al.* 2015) with the following options: k-list 33,55,77,99,127-min-contig-len 1000. All filtered reads were then mapped to the generated scaffolds using bowtie2 (Langmead and Salzberg 2012) with the –very-sensitive option. The mapping of the contigs was used to calculate the mean coverage of each contig. To ensure a reliable taxonomic assignation, only scaffolds longer than 5 kb were considered and submitted to metaGeneMark (v3.38) (Zhu *et al.* 2010) with the default metagenome-style model. The predicted genes were then compared with the RefSeq database (Li *et al.* 2021) using the Diamond version of blastP (v0.9.31) (Buchfink *et al.* 2015) with an *E*-value < 10^{-5}, identity percentage threshold of 35 and retaining only the best hits. The taxonomy of each scaffold was then inferred with a custom-made script applying a last common ancestor (LCA) method. At each rank, recursively, the taxonomy was conserved only if over half of the annotated genes presented the same keyword. Unclassified and ambiguous contigs were screened against RefSeq with Diamond blastX (–sensitive option and *E*-value < 10^{-5}), in order to detect exons. Non-overlapping best hits were used. The above taxonomical protocol was applied and only eukaryotic scaffolds were conserved. An alternative taxonomical annotation at the single read level was computed by Kraken (v2) (Wood *et al.* 2019) using its standard database. This annotation was found to be much less comprehensive and was not used further.

The diversity and sequencing effort of each sample was estimated by Nonpareil (v3.304) (Rodriguez-R *et al.* 2018) on forward reads over 35 bp using the *k*-mer algorithm.

Contig taxonomy and coverage data were used for beta-diversity estimation together with the *k*-mer frequency of reads. To do so, 15 bp *k*-mer frequency was retrieved for forward and re-

Table 1B. Cold surface soils samples analyzed in this study.

Sample	IGS code	Latitude (°)	Longitude (°)	Locality	Type
P2	C	54.54972	160.58194	Kamchatka, Kronotsky river bank	Vegetation free soil
P4	D	55.09861	160.34944	Kamchatka, Kizimen volcano	Tundra soil
P5	E	55.11500	159.96333	Kamchatka, Shapina river bank	Vegetation free soil

verse reads by Gerbil (v1.12) (Erbert *et al.* 2017). An in-house script was used to merge the produced files and compute the cosine distance between samples.

The Pfam domains (Nov. 2021 database version; Mistry *et al.* 2021) of all open reading frames (ORFs) were searched with InterProScan (v.5.39-77.0; Jones *et al.* 2014). The non-overlapping matches with *E*-value < 10^{-5} were retrieved with an in-house script.

β-Lactamase gene detection and analyses

ORFs with a best Diamond blastP match to a β-lactamase and β-lactamase-related proteins in the NCBI RefSeq database were retrieved and analyzed (Table S2, Supporting Information). Functional domains were inferred by InterProScan (v5) against Pfam, ProSitePatterns and by Batch CD-search (Marchler-Bauer and Bryant 2004) against the COG database (Galperin *et al.* 2021) (*E*-value < 10^{-5}) (Table S3, Supporting Information). ORFs at least 100 amino acids in length exhibiting a β-lactamase domain according to Pfam and COG (except for class D solely defined by COG domain) were kept for further analyses. We attempted to identify the β-lactamase-encoding contigs as potential plasmids using PlasClass (Pellow *et al.* 2020) and PlasFlow (Krawczyk *et al.* 2018). However, the results were found to be unreliable and not used further.

To normalize the β-lactamase count by a biologically relevant metric, we estimated the number of bacteria in our metagenomic dataset using 120 bacterial single copy genes from the Genome Taxonomy database using GTDB-Tk (Chaumeil *et al.* 2020) as reference (Table S4, Supporting Information).

For comparison, the abundance of β-lactamases was also computed in all of the terrestrial metagenomics datasets available from the IMG/M database from June 2021 (Chen *et al.* 2021). One thousand eight hundred two datasets were initially downloaded and then checked for contig length, coverage and quality. Four hundred thirty-four datasets from known biomes and of sufficient sizes (at least one contig over 10 kb and at least three bacteria equivalent sequenced) were retained and analyzed with the exact same protocol used to detect ORFs, β-lactamases and bacterial single-copy genes in our own dataset (Table S5, Supporting Information).

All statistical analyses were done using v4.1.1 of R (R Core Team 2021). Bray–Curtis dissimilarities were calculated with package vegan (v2.5-7) (Oksanen *et al.* 2020) and principal coordinate analysis performed with package stats4 (v4.1.1).

Results

DNA extraction yield and quality

Large variations were observed in the DNA contents of the different soil samples despite their similar macroscopic appearances (black/brownish fine silt and sandy compact soils), including between surface samples (taken from vegetation-free spots). Using the same extraction protocol, up to 75 times more DNA could be recovered from similarly dated ancient soils from different sites (Table S1, Supporting Information, for instance sample B vs P). A lesser range of variation (10 times) was observed between surface samples (samples C and D) already suggesting that the sample age was not the main cause of these differences. Interestingly, very similar total numbers of usable base pairs and good quality reads were determined from the 250 ng of purified DNA used for sequencing) (Table S1, Supporting Information). Except for sample B (sequenced as shorter reads on a different platform), we observed no correlation between the initial DNA content of the sample and the number of usable good quality reads (Table S1, Supporting Information), indicating that, once purified, all DNAs exhibited a similar quality.

Sample diversity

In contrast, the assembly of a similar number of sequenced reads from the different datasets resulted in a highly variable number of contigs (≥5 kb). As these values are linked to the complexity of the sample (species richness, number of different species; and their relative abundance, species evenness), this suggested that the samples exhibit globally different microbial population compositions and structures. As our samples are expected to be variable in richness and evenness while the number of reads is quite similar (except for sample B), the sequencing effort is expected to be different for each sample. This was confirmed by the Nonpareil tool that calculates the sequencing effort and an alternative to the Shannon diversity index computed from the *k*-mer diversity of reads (Rodriguez-R *et al.* 2018). Prior to any taxonomical annotation we could thus estimate the sample diversity as well as the estimated sample coverage (i.e. the proportion of the sample diversity actually sequenced) that ranges from 0.39 (sample E) to 0.9 (sample L) (Fig. S2, Supporting Information). Given the same number of reads for two samples, a larger coverage corresponds to a lesser diversity. According to Fig. 1, the richest is sample E and the poorest is sample L. There are large differences between samples, including among the Kamchatka surface samples that appear the most diverse. The Nonpareil diversity values do not significantly correlate with the amounts of DNA recovered from one gram of soil (Table S1, Supporting Information). In contrast and as expected, the diversity values negatively correlate with the median coverage of the contigs in each sample (Pearson correlation of −0.75, $P < 0.005$).

When focusing on a given borehole, we found that the diversity appeared to increase with sample depth in a given borehole (Fig. 1). However, diversity values were quite variable for samples taken at similar depths across different boreholes, in line with the lack of a consistent relationship between the diversity and the age (from radiocarbon dating) of the sample (Table 1A).

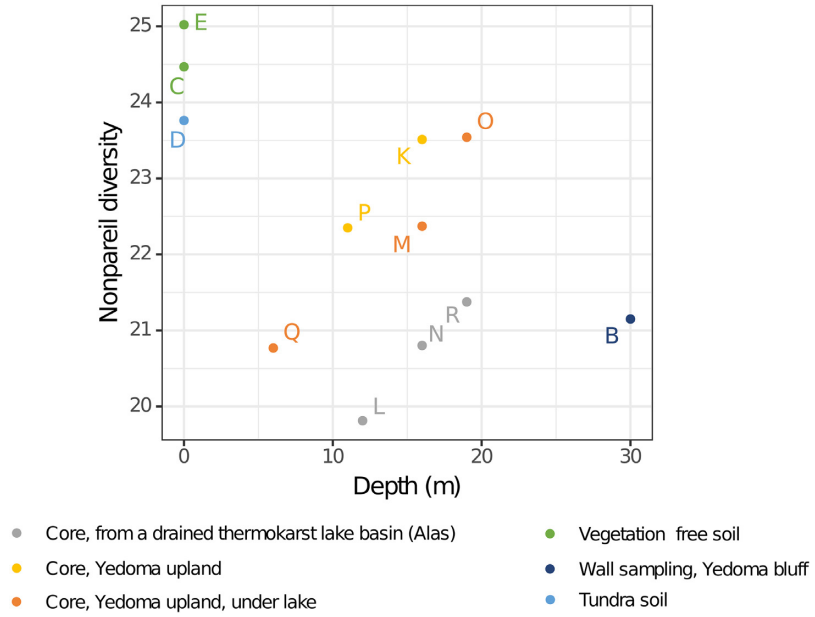

Figure 1. Nonpareil diversity compared with sample depth. Nonpareil diversity was computed on reads over 35 bp in size. This metric is correlated to the Shannon index (Rodriguez-R *et al.* 2018).

Global species content

As depicted in Fig. S3A (Supporting Information), bacteria account for over 90% of the estimated abundance in most samples and 87% in average but they represent 95% of contigs (>5 kb). This is true for all modern cryosols and most ancient permafrost layers despite their different origins and DNA contents. Even higher proportions of bacterial sequences were computed by Kraken2 (Wood *et al.* 2019) (Fig. S3B, Supporting Information). However, we found this method to be less sensitive and was not used further in this study. Four samples, L, N, R and Q, appear to be poorer in bacterial sequences (≤77%) due to their higher proportions of unclassified and Archaea sequences (Fig. S3A, Supporting Information).

In all samples combined, Archaea represent 3.2% of the abundance but only 1.8% of the total contig number. In most samples, they are very rare except in sample Q (unfrozen talik) and three samples from different depths in the same borehole (L, R, N: respectively, 12, 16 and 19 m) where they represent 6.5–11.3% of the total abundance (Fig. S3A, Supporting Information). In the samples where they are in sizable amounts, the archaeal population is dominated by Bathyarchaeota (a recently described clade of methanogens belonging to the TACK phylum) (Evans *et al.* 2015), Nitrososphaeria (a class of chemolithoautotrophic ammonia oxidizing Archaea belonging to the Thaumarchaea and globally distributed in soils) (Hatzenpichler 2012) and Methanomicrobia (a class of methanogens belonging to the Euryarchaeota and previously noticed to be enriched in the sediment microbiomes associated with the rhizosphere of macrophytes) (Behera *et al.* 2020) (Fig. S4, Supporting Information).

Eukaryotic DNA represents <1% of contigs and 0.5% of the total abundance, except for two subsurface samples from the same borehole (R, N) with 1% and 2% of eukaryotes respectively (Fig. S3, Supporting Information). The Viridiplantae (land plants and green algae) and Unikont (mostly Fungi, Metazoan and Amoebozoa) constitute most of the represented clades except for samples E and O (Fig. S5, Supporting Information). The latter only exhibits 68 total eukaryotic contigs, 21 of which were attributed to the Bacillariophyceae (the class of diatoms). With 30 other contigs most likely originating from green algae, this suggests that this 40 000-year-old layer originated from a lake bottom deposit (Table 1A) (Fig. S5, Supporting Information).

Finally, according to the LCA assignment method, viruses were predicted to only represent 0.4% of the global coverage of all samples (Fig. S3, Supporting Information), with one sample (R) standing out by reaching 2.3%. Such low representation prompted us to reanalyze our datasets using VirSorter, a leading software tool specifically dedicated to the identification of viruses in metagenomics data (Roux *et al.* 2015). This new protocol increased the average abundance of virus to 2.68%, with only a handful of samples exhibiting values higher than 2%, including the R sample at 8.6% (Fig. S6, Supporting Information). The newly annotated viral (mostly phage) contigs originated from contigs attributed to Bacteria by the LCA method (Fig. S6B, Supporting Information). The family distributions between theses samples is quite variable but, in general, the most abundant viruses are Mimiviridae, Pithoviridae (the prototype of which was previously isolated from ancient permafrost; Legendre *et al.* 2014), Caudovirales, Phycodnaviridae and Ascoviridae, all families known to gather giant or large DNA viruses. A detailed analysis of these virus populations will be published elsewhere.

A high diversity that points out syngenetically trapped communities

The global community analysis and an analysis of the *k*-mer frequencies of reads in each dataset do not consistently put samples from the same borehole close to each other (L-N-R, P-K, Q-M-O) (Fig. S7, Supporting Information). These analyses do not allow a meaningful clustering of samples originating from the different boreholes, or even distinguish surface samples from permafrost samples (Fig. S7, Supporting Information).

Focus on the bacterial populations

The relative abundances of phyla representing >0.5% of the total sum of coverages are depicted in Fig. 2. Actinobacteria and Proteobacteria are the two most abundant phyla. Although

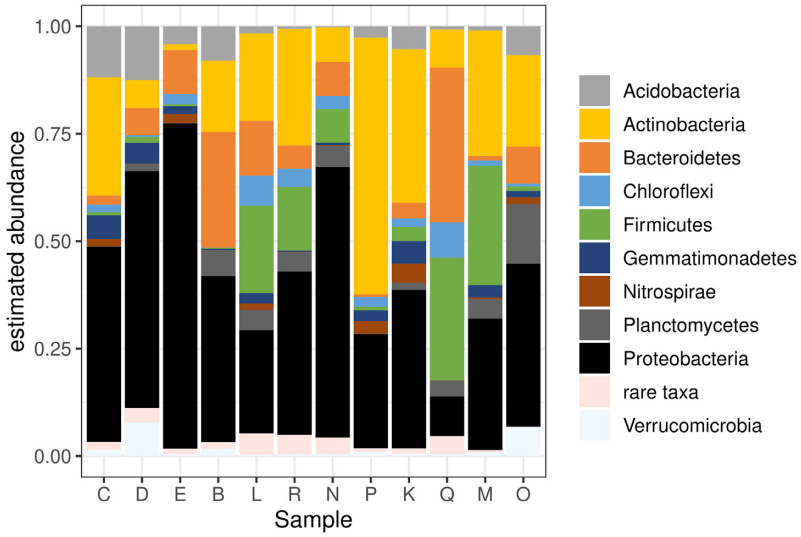

Figure 2. Bacterial phyla abundances across samples. The abundance is estimated as the fraction of the total sum of coverages.

Proteobacteria appear globally dominant (40% of the estimated abundance and over 75% in sample E), it is not always the most abundant in every sample. The most extreme cases include sample P in which Actinobacteria are largely dominant, and sample Q in which Bacteroidetes is the leading phylum. The relative abundance of Firmicutes, Planctomycetes, Acidobacteria and Verrumicrobia varies strongly across samples from up to 28%, 14%, 13% and 7.8%, respectively, almost down to zero. None of these characteristics appears to correlate with the geographic location, depth or age of the samples.

Within a given phylum, such as the Proteobacteria (Fig. S8, Supporting Information), very distinct distributions across classes could be seen between different modern cryosols (C, D, E) or ancient permafrost layers either from the same borehole (L-R-N, Q-M-O) or from similar depth (16 m: M and N; 19 m: R and O). Species from the Alpha, Beta and Delta/Epsilon divisions constitute the bulk of the Proteobacteria populations, although with strong variations. For instance, Betaproteobacteria are quasi absent from the R and N samples, as are Alphaproteobacteria from sample E.

At lower taxonomical ranks the distributions become scattered, as well as less informative given the smaller proportion of non-ambiguously annotated contigs (66.3% at the phylum level, 43.6% at the order level, but only 16.4% at the genus level). The most abundant order overall is Rhizobiales that is present in all samples (Fig. 3). Unclassified Actinobacteria are the second most abundant group suggesting that the cryosol Actinobacteria are very different from their cultured relatives. In samples taken individually, Rhizobiales is the most abundant order in samples N and R where they represent 53% and 23% of the bacterial abundance, respectively. Burkholderiales are the most abundant order in samples E, D and O and Clostridiales are the most represented in samples Q and O. In the other samples, unclassified Actinobacteria represent the major group. In some samples (C, D, E, B, L, O), the high abundance of Proteobacteria is divided between multiples orders in the phylum rather than due to a predominant one. In samples R, N, P, K, M on the other hand, Proteobacteria are mostly represented by Rhizobiales (Fig. 3). Some of the orders found also in deep samples are known to be aerobic. This is the case for Vicinamibacteraceae (Huber and Overmann 2018) and Gaiellales (Albuquerque and da Costa 2014) but these together constitute only 2.4% of the bacterial abundance. Most groups present in these cryosols are thus facultative anaerobes and chemotrophs.

The analysis suggests a lot of variability across samples, making it illusory to expect that a tractable number of bacterial orders could be used to classify the various types of cryosol and permafrost into recurrent taxonomical types. While surface samples and most deep samples form separate clusters (Fig. 3), we should be careful in drawing conclusions as the surface samples come from separate locations. When samples from the same borehole do cluster together (L-N-R and P-K), their dominant bacterial communities still exhibit some variations (Fig. 3). Interestingly, the large distance separating sample Q from the same borehole samples M and O is consistent with its unfrozen state as part of an under lake talik (Table 1A). Statistical analysis of similarities (ANOSIM) confirms these observations (Fig. S7, Supporting Information) showing that the distances between samples from the same borehole are not significantly lower than the distances between samples from different boreholes ($R = 0.388$, $P < 0.069$).

As a complimentary attempt to interpret the microbiome differences between samples, we screened the Pfam database, identified a total of 9481 motifs associated with specific protein families and/or molecular functions, and computed their distribution across samples. We expected that this type of analysis could reveal ecological/environmental constraints shared by various samples in spite of their different microbial communities. A list of the 50 most abundant Pfam motifs globally found in our dataset is given in Table S6 (Supporting Information). This list includes 7 of the 20 most abundant motifs found in the whole Pfam database, supporting the validity of our annotation protocol. Figure 4A presents a Bray–Curtis dissimilarity map computed from the Pfam counts in the samples. This map confirmed the functional similarity of the L-N-R samples, of the P-K samples (with the insertion of M from a different borehole, but from a similar depth and age as K) and to a lesser extent of two surface soil samples (D, E). The rest of the samples exhibited unrelated functional patterns, with the unfrozen talik-derived sample Q as the most dissimilar.

Table S6 (Supporting Information) also indicated the presence of three Pfam motifs, found in proteins involved in bacterial antibiotic resistance: TetR-like transcription regulators (PF00440, ranked 29th/9481), metallo-β-lactamases (PF00753, ranked 44th)

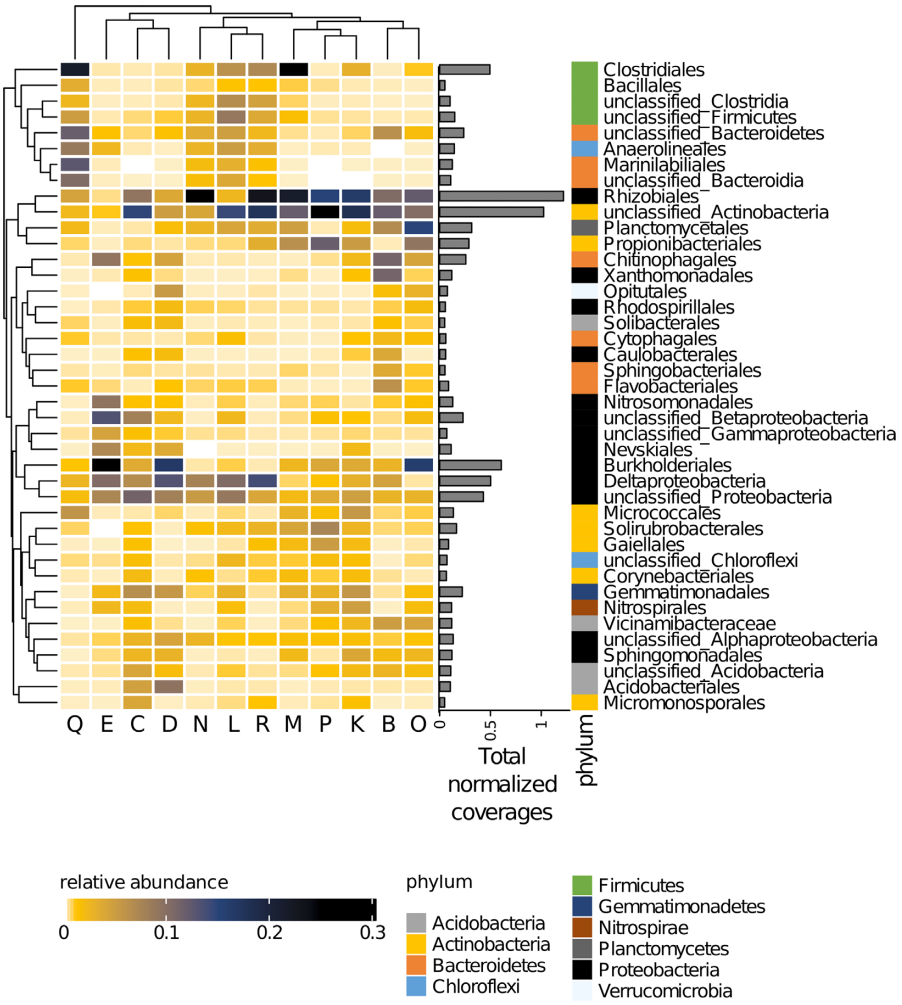

Figure 3. Bacterial abundance across samples. Samples and most abundant bacterial orders (over 0.5% of the total normalized coverage) were clustered using the Bray–Curtis distance. The abundance is estimated by the sum of coverages (normalized to 1 in each sample).

and glyoxalases (PF00903, ranked 46th). This prompted us to focus our next analysis on Pfam motifs putatively associated with virulence factors, the results of which are presented as a heat map in Fig. 4B. In this graph, three different protein families associated with bacterial antibiotic resistance appear to be largely dominant. The top one corresponds to a large family of enzymes catalyzing different reactions, including glyoxalases, dioxygenases and bleomycin/fosfomycin resistance proteins. Unfortunately, this Pfam motif is not specific enough to distinguish the minority of proteins truly involved in antibiotic inactivation from the rest of the family. Fortunately, the next most abundant functional families (β-lactamases and metallo-β-lactamases) are much better defined and for the most part correspond to enzymes active against various types of β-lactam antibiotics. Their distribution is further analyzed in the next section. Finally, the less abundant TetR (tetracycline resistance) family in the heat map is defined by an ambiguous motif found in a large diversity of DNA-binding proteins most of which are not involved in antibiotic resistance.

Bacteria carrying β-lactamase genes

Given the central role of extended-spectrum β-lactamases (ESBL) in the spread of multiple bacterial antibiotic resistances (Cantón *et al.* 2012, Potron *et al.* 2015, Wright 2019) we specifically investigated the presence and diversity of these proteins/enzymes in our dataset. Although the ancient microorganisms trapped in the pristine permafrost layers studied here have little probability to come in direct contact with human beings, the risk remains that they might contribute to new β-lactam resistances once in contact with modern bacterial species.

A total of 1071 β-lactamase genes were detected by diamond blastP against RefSeq and confirmed by a domain search (Fig. S9, Supporting Information). Among them, 1066 were found in bacterial contigs, 4 in an archaeal contigs and 1 in an ambiguous contig. Figure 5 presents the numbers, origins and classes of bacterial β-lactamases. Although these numbers are small compared with the total number of contigs (Table S1, Supporting Information), they nevertheless suggest that, in average, most (87%) of the sampled Bacteria harbor one copy of a β-lactamase gene, once normalized by the frequency of 120 known single copy genes detected in each sample (Fig. 5; Table S4, Supporting Information).

The normalized β-lactamase copy-number ranges from 0.39 [confidence interval: 0.22–0.55] per bacteria (sample N) to 1.75 [confidence interval: 1.54–1.97] (sample K) (Fig. 5). We noticed that the lowest prevalence of β-lactamase-encoding bacteria corresponds to samples from the same boreholes (L-R-N and Q-M), suggesting an actual ecological constraint. In most cases, we identified a single β-lactamase copy per contig, except for 17 contigs carrying several occurrences. In 8 of these, two β-lactamase genes

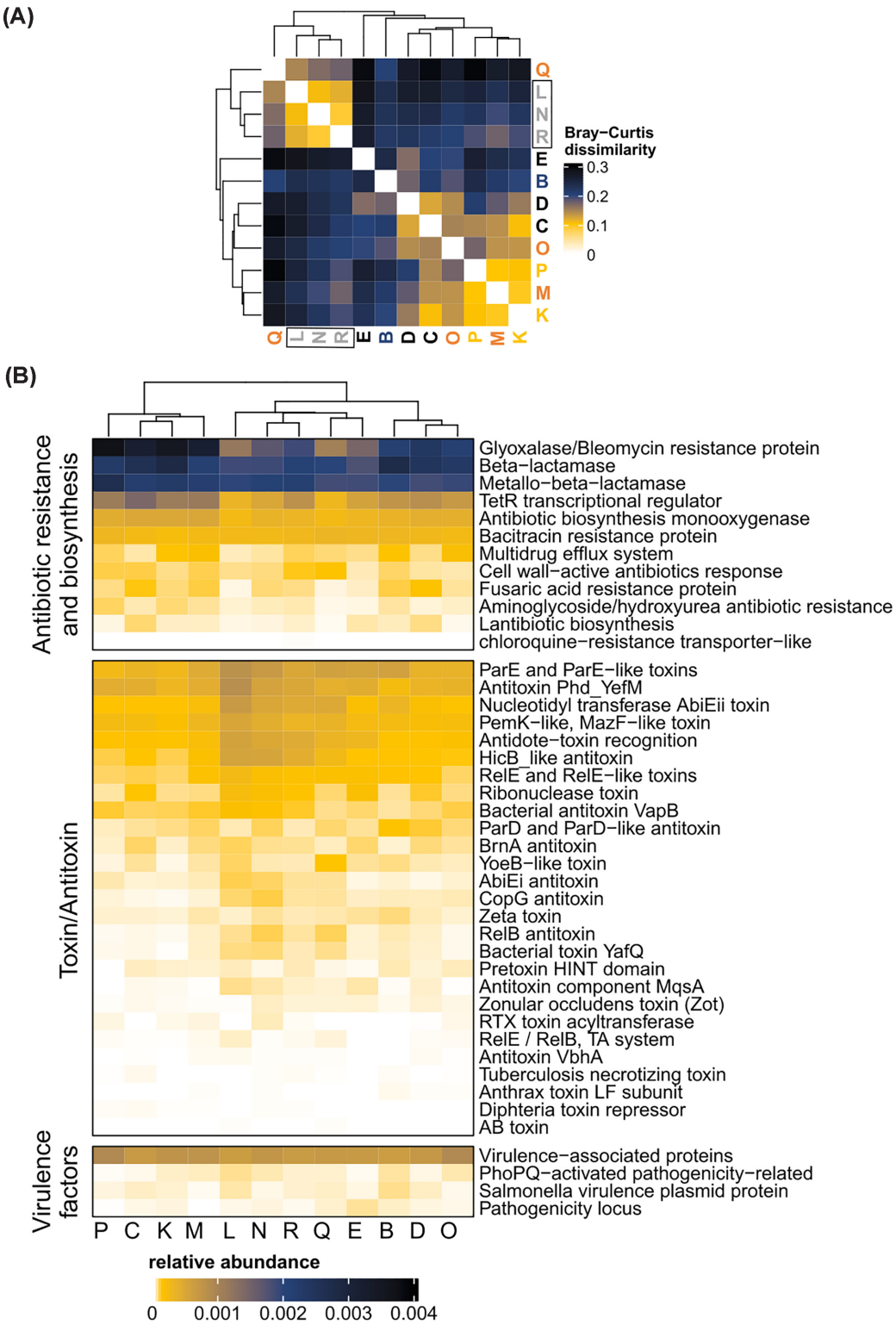

Figure 4. Analysis of the functional similarity between samples. **(A)** Bray–Curtis dissimilarity matrix computed from all Pfam counts in bacterial contigs. **(B)** Heat map of the abundance of Pfam motifs potentially associated with bacterial virulence factors.

were located next to each other and in 14 contigs, the two or three β-lactamase genes were found <10 ORFs away from each other.

Despite the large fluctuation in the average β-lactamase copy number, the relative proportions of the four enzyme classes remain fairly stable across samples. Class C (52–83%) was always the most represented, followed by class B (14% in average), then Class A and Class D (10%) (Fig. 5). All classes were present in every sample except for sample E (no class A β-lactamase).

Finally, we investigated the distribution of β-lactamase genes among the various bacterial phyla, as one given group of bacte-

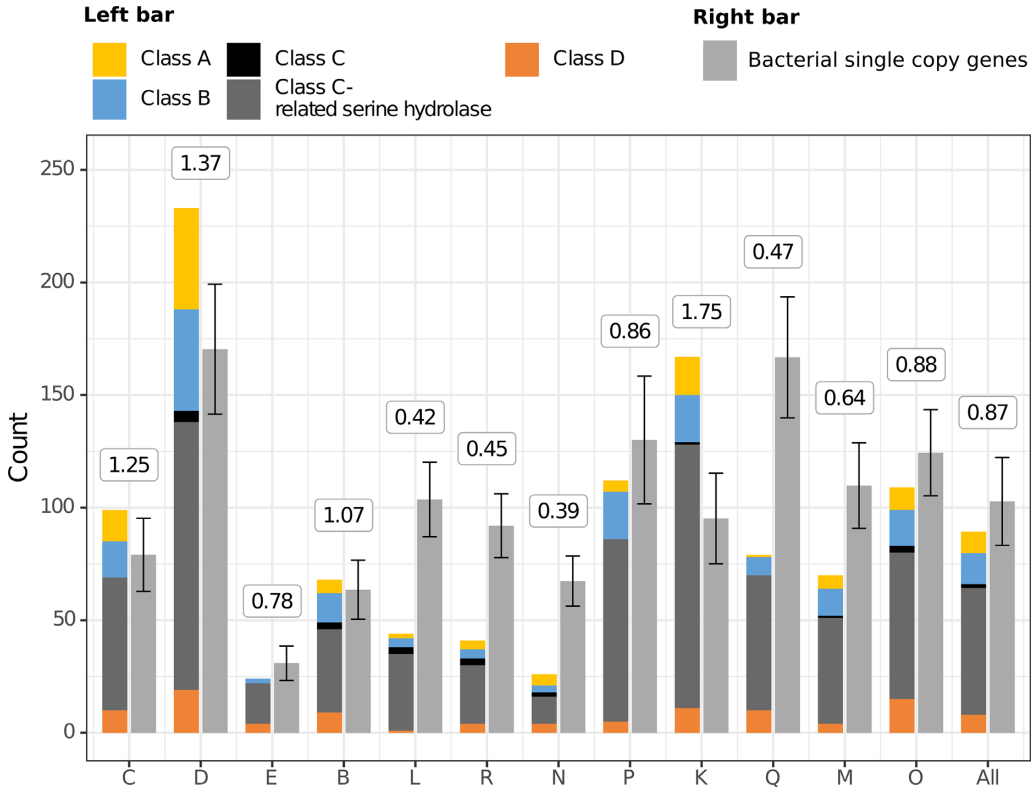

Figure 5. Bacterial β-lactamase classes and copy numbers. The distribution of the four main Ambler's classes of β-lactamases in the various samples is shown by the left bars. 'Class C-related serine hydrolase' stands for proteins with a β-lactamase AmpC domain and whose best diamond blastP matches were annotated 'β-lactamase-related serine hydrolase'. The number of bacterial single copy genes is shown by the adjacent gray right bar (mean count out of 120 reference single copy genes, with error bars corresponding to standard deviation). The boxes on top of each barplot indicate the average β-lactamase copy number per bacterial cell.

ria could concentrate most of β-lactamase genes. We thus simply compared the relative abundance of each bacterial phylum within the set of all cognate contigs with that of those exhibiting a β-lactamase gene. Bacteroidetes appeared to be the most enriched group in β-lactamase genes, while Planctomycetes exhibited the lowest proportion. Although the two distributions (Fig. 6) are different, they remain significantly correlated (Pearson coefficient > 0.8) pointing out that the occurrences of β-lactamase genes in these cryosols were shared among the diverse phyla constituting these microbiomes.

Comparing β-lactamase gene occurrences in the permafrost versus other soils types

To compare the β-lactamase gene proportion found in the above Russian cryosols to other types of soils, we extended our analysis to over 1500 terrestrial metagenomic assemblies publicly available from the JGI. As soil communities are usually very complex (e.g. Rodriguez-R *et al.* 2018), their assembly into large contigs is usually challenging (Alteio *et al.* 2020). Accordingly, only 434 datasets were found suitable to be further analyzed by our protocol. If the average number of β-lactamase genes per bacteria was again found to be close to one, it was quite variable (mean: 1.09 ± 0.77) (Fig. 7). Interestingly, if our samples did not stand out, other permafrost samples from Alaska and north Sweden ranked high in the list (Fig. 7). The samples with most occurrences of β-lactamase genes per bacteria were agricultural soils followed by riparian soils. There is significantly more β-lactamase copy number in agricultural samples than in other terrestrial soils except for the JGI permafrost samples and riparian soils (Fig. 7). Bog and forest soils also exhibited β-lactamase gene copy numbers close to one. The environments exhibiting the smallest β-lactamase gene copy numbers are shale carbon reservoirs, uranium contaminated soils and mire (a kind of wetland) (Fig. 7).

Discussion

A variable DNA content

We found the quantities of DNA recovered from the various cryosols to be highly variable, from >10 μg/g to 0.02 μg/g, thus a factor of 500 (Table S1, Supporting Information). Most of the low yielding samples are from ancient (subsurface) permafrost layers, albeit without a strict correlation with age. The Stanchikovsky Yar sample (B) exhibited a particularly high DNA content. However, once purified, the recovered DNA exhibited similar sequencing yields (i.e. number of reads/ng) and assembled into contigs of sizes similar to that of other samples (Table S1, Supporting Information). This suggests that the variation in DNA content is not due to its degradation, but reflects real differences in the global amount of microorganisms per gram of sample. Such a result is significant because it supports the hypothesis that genes encoded by the ancient DNA could be recycled within contemporary microorganisms through transformation.

An unexpected increase of diversity with sample depth

The nonpareil diversity (a good correlate of α-diversity and of the Shannon index) (Rodriguez-R *et al.* 2018) was found to increase with depth for each of the three boreholes. This is in

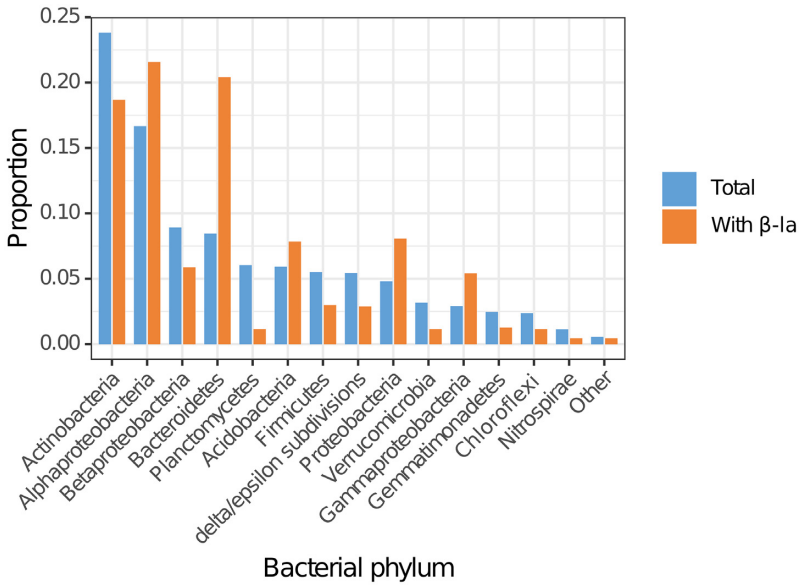

Figure 6. Side-by-side comparison of the proportion of bacterial phyla in all samples with the distribution of β-lactamase carrying bacteria. These proportions are globally correlated (Pearson correlation coefficient > 0.8), indicating that β-lactamase genes do not originate from a single dominant phylum. Bacteroidetes appeared to be the most enriched phylum in β-lactamases.

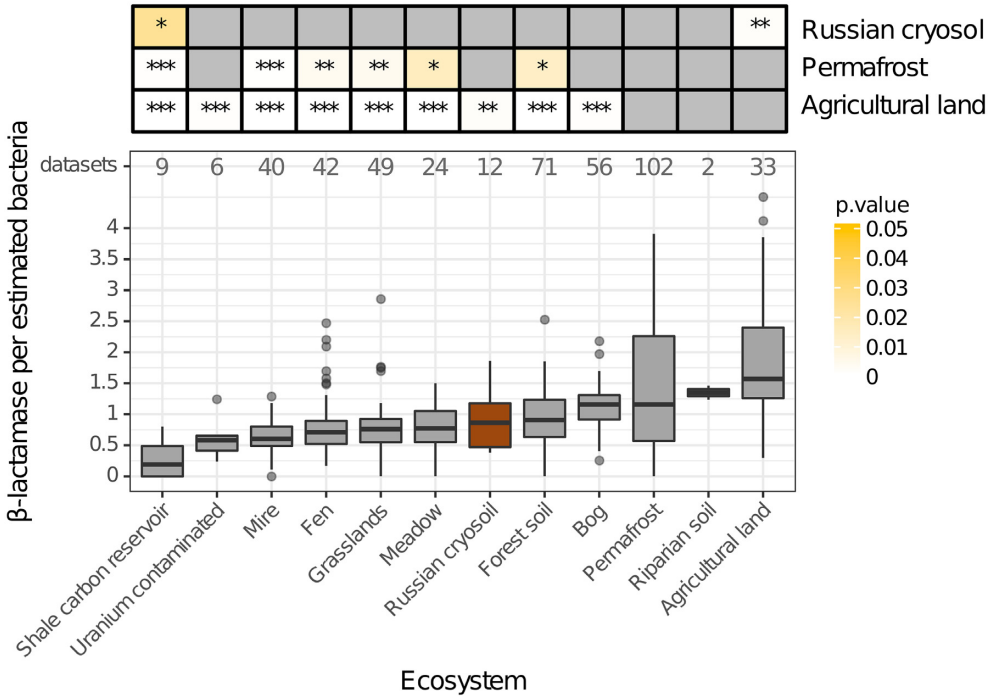

Figure 7. β-Lactamase gene copy numbers in available terrestrial metagenomes. The values are computed as the ratios between the β-lactamase gene counts and the average number of single copy genes identified in each sample by GTDB-Tk. Boxes are ranked according to the median of the β-lactamase gene copy number. Only datasets with three or more estimated entire bacteria sequenced were included. The brown box corresponds to the 12 new metagenomes analyzed in this work. The P-values of the pairwise comparison (Wilcoxon test) between our samples, agricultural lands and permafrost to other datasets are shown in the above heat map. Empty gray boxes correspond to nonsignificant differences.

contrast to the trend previously reported for culturable permafrost microorganisms (Gilichinsky *et al.* 1992). As argued by Abramov *et al.* (2021), microorganisms trapped in permafrost do not multiply and, because of gradual death, their diversity is expected to decrease over time. They claimed that, although growth as subzero temperature remains possible for some microorganisms (for instance, Panikov and Sizova 2007, Mykytczuk *et al.* 2013), the nutrient level and diffusion is too low to sustain growth as seen in the laboratory. The trend exhibited in our data between age and diversity is thus probably due to actual variations in diversity between syngenetically trapped microbiomes.

Scarcity of eukaryotes

A striking finding of our taxonomic analyses is the very low representation of eukaryotic organisms, only present in trace amounts in all our samples except for two deep samples (R, N) (Figs S3 and S5, Supporting Information). This result is unexpected given the fact that the warm spring–summer period sees the growth of a dense vegetation cover, accompanied by a variety of mosses and fungi, and a proliferation of insects, which should leave lasting traces in the metagenome of permafrost, once buried. Furthermore, the most abundant bacteria we found belong to the order Rhizobiales, which are often in symbiosis with plants. The Siberian permafrost harbored also members of Burkholderiales (6% in average, up to 27%), Chitinophagales (3% and up to 9%) and Cytophagales (0.7%), all of which are enriched in fungal–bacterial networks (Bonito *et al.* 2019). A previous culture-based study of Siberian, Canadian and Antarctic permafrosts estimated the count of fungal cells around 10^2 per gram of dry soil, distributed among 49 species (Ivanushkina *et al.* 2005). As fungi are in form of spores in permafrost (Vorobyova *et al.* 2001), part of their diversity might remain unseen because of difficulties to extract their DNA. The proportion of eukaryotic cells might also be much lower than that of bacteria. In such case, their sequencing coverage might be insufficient to assemble contigs of a least 5 kb in size. The larger size of eukaryotic genomes, the presence of introns and their lower coding density due to large intergenic regions also strongly decrease the likelihood of their identification by similarity searches in protein databases. However, *k*-mer-based read classification (Fig. S3B, Supporting Information) is consistent with an overall low proportion of eukaryotes. The remaining possibility is thus that eukaryotic DNA is rapidly recycled into microbial DNA during the seasonal decomposition process that takes place in the superficial active layer before being perennially frozen in deeper permafrost.

Variable bacterial community composition

As expected, bacteria constituted most of the DNA sources (Fig. S3, Supporting Information). Compared with barcoding methods, metagenomics allows an exploration of the sample diversity that is, in principle, only limited by a threshold of minimal abundance for a given species. This threshold results from three main factors: (i) the proportion of cells (or viral particles) of a given species (mostly unicellular) in each sample, (ii) the efficiency with which their DNA can be extracted (e.g. sporulating vs nonsporulating species) and (iii) our capacity to identify them by sequence similarity searches in reference databases. The last factor involves variations in protein-coding density (high in viruses and prokaryotes, much lower in most eukaryotes except for some fungi and protozoans), the fraction of the database corresponding to known members of the various phyla (high for Bacteria, much lower for viruses and Archaea), the level of sequence divergence among homologues within different domains (very high in viruses, lower in eukaryotes). Thus, even for a similar coverage (i.e. number of redundant reads per genome position) in a metagenomics mixture, bacteria are more likely to be identified than archaeal and eukaryotic organisms, viruses being the least likely to be identified (i.e. taxonomically unclassified). The comparisons of organism frequencies across domains are thus much less reliable than within a given domain. Yet, the predominance of the bacterial domain is unlikely to be due to a bioinformatic/annotation bias. To estimate the abundance of a given group we used the sum of coverages of the contigs attributed to the group. In theory, a better estimation would have been the number of attributed reads relative to the genome size (Nayfach and Pollard 2016). In the absence of information on the genome sizes, the sum of coverages thus remains a good approximation.

Globally, the phyla found to dominate the microbiomes in our cryosol and permafrost samples do not differ from those already reported as most abundant in previous analyses of temperate soils (Janssen 2006) or permafrosts (Tripathi *et al.* 2019, Xue *et al.* 2019). Those are Proteobacteria (40%), Actinobacteria (22%), Bacteroidetes (10%), Firmicutes (9%), Acidobacteria (5%), Planctomycetes (4%) and Chloroflexi (3%) (Fig. 2). However, even the most common phyla do exhibit very different proportions between samples: 76–9% for Proteobacteria and 59.7–1.4% for Actinobacteria (Fig. 2). We noticed that the same phyla dominate both in the surface cryosols and the permafrost sample except for members of the Firmicutes that are mostly found below the surface (<1.3% at the surface, up to 28.5% in permafrost). In contrast to a previous study (Tripathi *et al.* 2019), we did not detect a significant amount of Crenarchaeota. On the opposite, we detected Bathyarchaeota, absent from this same study (Fig. S4, Supporting Information).

The high taxonomical variability depicted here confirms that no significant mixing occurred between the distinct microbiomes of the permafrost layers, in particular for the Q, M, O samples extracted from the same borehole (Figs 2 and 4). The weak clustering of samples L, N, R and samples P, K using the Bray–Curtis dissimilarity might solely reflect the influence of the soil type or environmental conditions shaping bacterial communities.

Although dead cells would result in altered DNA and thus a lower efficiency of sequencing and assembling large enough contigs, it remains possible that most of the bacteria identified in our study might be dead. A previous metagenomic study of ancient Alaskan permafrost estimated that only 15–18% of cells were alive. However, the DNA depletion to remove dead cells did not significantly changed the relative proportion of bacterial groups (from phylum to family) or the phylogenetic diversity (Burkert *et al.* 2019).

There is a growing concern that taxonomy alone is insufficient to interpret the complexity of soil processes (Baldrian 2019) given the already high diversity of metabolisms and lifestyles encountered in a given bacterial group. A good example of such a difficulty is the presence of Anaerolineales species in all our samples (Fig. 3). Although their anaerobic metabolism is consistent with life in deep permafrost, all known isolates exhibit optimal growth temperature in the 37–55°C range (Sekiguchi *et al.* 2003, Yamada *et al.* 2006) or come from hot environments (Konishi *et al.* 2012). On the other hand, these detected Anaerolineales species appear to cohabit with close relatives to known psychrophilic bacteria such as *Devosia psychrophila* and *Dyadobacter psychrophilus*, the presence of which is not unexpected in cryosols.

Thawing permafrost as a source of ARGs

Our study indicated the presence of the four main classes of β-lactamases (Ambler 1980), the evolution of which can rapidly lead to ESBL in the context of strong antibiotic selection in clinical settings. Class A EBSL includes cephalosporinases and six types of dreaded carbapenemases (Sawa *et al.* 2020). Class B β-lactamases are metalloenzymes using zinc at their active center compared with a serine residue for other classes. They hydrolyze carbapenems and degrade all β-lactam agents except monobactams (Sawa *et al.* 2020). Class C β-lactamases derives from the ampC gene carried on the genome of many Enterobacteriaceae. Variants of these enzymes are known to reduce sensitivity to carbapenems (Sawa

et al. 2020). Finally, class D β-lactamases, also known as oxacillinases (OXAs) evolved from degrading an extended spectrum of cephalosporins to hydrolyze carbapenems (Sawa *et al.* 2020). However, from their sequences alone (except for near 100% identical residues) it is impossible to predict the substrate range and the potential clinical risk associated with the diverse β-lactamases identified in our dataset.

The presence of antibiotic-resistant bacteria carrying various types of β-lactamases has been previously reported in pristine Arctic and Antarctic surface cryosols and ancient permafrost (Allen *et al.* 2009, D'Costa *et al.* 2011, Petrova *et al.* 2014, Perron *et al.* 2015, Kashuba *et al.* 2017, Van Goethem *et al.* 2018, Haan and Drown 2021). However, these previous studies did not evaluate the proportion of bacteria carrying a β-lactamase gene or exhibiting a β-lactam resistance in the global population of soil bacteria, most of which are unculturable (or eventually dead).

The number of identified β-lactamase genes was divided by the estimated number of distinct bacteria to obtain a normalized gene count. This original approach provides a biologically relevant quantification that can be generalized to target other enzymes and detect specifically enhanced metabolic features in prokaryotic communities. The usual approach normalizes the ORF count by the number of reads, and is less biologically relevant than the ORF count per bacteria.

We found few previous studies against which comparing our results about the presence of β-lactamases in a quantitative manner. Almost all investigations of environmental 'resistomes' were performed on populations of culturable bacteria sampled near the surface and selected for their antibiotic resistance (e.g. Haan and Drown 2021). A few others were restricted to a given bacterial species (e.g. *Escherichia coli*) (Pormohammad *et al.* 2019). We found one study analyzing the proportion of β-lactam-resistant bacteria within different agricultural soils, untreated or amended with manure (Udikovic-Kolic *et al.* 2014). This proportion was found to vary from 0.67% to 7.4% in untreated versus manure-amended soils. Most of these resistant bacteria presumably encoded a β-lactamase, thus in a much lesser proportion than found in our cryosol samples. Another metagenomic analysis of the distribution of ARGs in 17 pristine Antarctic surface soils estimated the relative frequency of ARG to all genes in the [1.3–4.4 10^{-5}] range, which approximately correspond to 17% of soil bacterial encoding at least one ARG (Van Goethem *et al.* 2018) (assuming an average content of 3850 genes per bacterium; diCenzo and Finan 2017). Moreover, this fraction includes all ARG (e.g. multidrug resistance efflux pump, aminoglycoside acetyltransferase/nucleotidylyltransferase, aminocoumarin resistant alanyl-tRNA synthetase, etc) of which β-lactamases only constitute a small proportion.

The high proportion of β-lactamase-encoding bacteria found in our 12 cryosol samples initially appeared significantly greater than previously reported in the few publications concerning other pristine environments, undisturbed by anthropogenic activities. However, following the analysis of 434 supplementary datasets, the proportion of β-lactamases in our 12 samples came out near the average of its value for a variety of other soils, and for instance twice less than in some American agricultural lands (Fig. 7). Other permafrost samples turned out to be richer in β-lactamase than ours, some even overlapping with the level of cultivated soils (Fig. 7). The comparison of β-lactamase copy numbers between biomes revealed statistically significant differences (Fig. 7). Thus, β-lactamases are not distributed evenly among diverse environments under the influence of unknown factors that remain to be identified in future studies.

The function of β-lactamases in soil microbial communities is not clear, but our results might help building new hypotheses. Several soil microorganisms produce antibiotic β-lactams, including members of the Actinobacteria (e.g. *Streptomyces* species) (Ogawara 2016). It turns out that these bacteria are among the most abundant in the cryosol microbiomes (Fig. 2). Among eukaryotes, fungi (also abundant in our samples, Fig. S5, Supporting Information) can also produce many complex β-lactam-related compounds. They are the original source of two foundational β-lactam antibiotics: penicillin and cephalosporin (Brakhage *et al.* 2009). The high β-lactamase abundance might thus result from the selective advantage they confer to their bearers in the middle of a biochemical war between soil microbes.

Alternatively, β-lactamase may also play a role in a less violent scenario, outside of the simplistic 'war' paradigm. Below the minimal inhibitory concentration, antibiotics may modulate bacterial gene expression, thus acting as signaling molecules rather than lethal weapons. In such context, β-lactamases might interfere with quorum sensing (Yim *et al.* 2007). Another possibility is that some of these genes encode a broader function (such as carboxylesterases or DD-carboxypeptidases active on β-lactams (Higgins *et al.* 2001, Nan *et al.* 2019, Pandey *et al.* 2020) providing a fitness gain in addition to protecting against soil antibiotics).

Unexpectedly, this study revealed a new potential danger, due to the abundance of β-lactamases found encoded in a large phyletic diversity of cryosol bacteria. The DNA from these permafrost bacteria, dead or alive, thus constitutes an immense reservoir of historical ARGs. Their transfer to contemporary pathogenic bacterial in a clinical setting might further contribute to the antibiotic resistance crisis, now considered one of the biggest public health challenges of our time.

In a context of global warming, ancient permafrost becomes an antibiotic resistance flavored ice cream ready to be consumed without moderation by all passing bacteria.

Acknowledgements

We are deeply indebted to our volunteer collaborator, Alexander Morawitz, for collecting the Kamchatka soil samples. We thank Dr Audrey Lartigue for technical advice about DNA extraction and acknowledge the computing support of the PACA BioInfo platform. We thank M. Ulrich (DFG project #UL426/1-1) and P. Konstantinov for helping with fieldwork at the Yukechi site, as well as the Alfred Wegener Institute and Melnikov Permafrost Institute logistics for field support and sample acquisition.

Supplementary data

Supplementary data are available at *FEMSML* online.

Contributions

J-MC and CA initiated the project and designed the study; SR and SS performed the bioinformatic analyses; J-MC wrote the initial version of the manuscript; EC-F processed the sample and performed DNA extraction; KL supervised DNA sequencing; and GG, JS and ANF collected the deep core samples from the Yukechi site. All authors contributed to the writing of the final manuscript.

Data availability

The data underlying this article are available in the EMBL-EBI database at http://www.ebi.ac.uk/, and can be accessed with

the study number PRJEB47746 (secondary accession: ERP132049), and samples accession numbers: ERS7649018, ERS7649019, ERS7649020, ERS7649021, ERS7649022, ERS7649023, ERS7649024, ERS7649025, ERS7649026, ERS7649027 and ERS7649028.

Funding

This work was supported by the Agence Nationale de la Recherche grant (ANR-10-INBS-09-08) to J-MC and the CNRS Projets de Recherche Conjoints (PRC) grant (PRC1484-2018) to CA. EC-F was supported by a PhD grant (DGA/DS/MRIS #2017 60 0004). GG and JS were funded by a European Research Council starting grant (PETA-CARB, #338335) and the Helmholtz Association of German Research Centres (HGF) Impulse and Networking Fund (ERC-0013).

Conflict of interest statement. None declared.

References

Abramov A, Vishnivetskaya T, Rivkina E. Are permafrost microorganisms as old as permafrost? *FEMS Microbiol Ecol* 2021;**97**: fiaa260.

Alberti A, Poulain J, Engelen S *et al.* Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci Data* 2017;**4**:170093.

Albuquerque L, da Costa MS. The family Gaiellaceae. In: Rosenberg E *et al.* (eds). *The Prokaryotes: Actinobacteria*. Berlin, Heidelberg: Springer, 2014, 357–60.

Allen HK, Moe LA, Rodbumrer J *et al.* Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* 2009;**3**:243–51.

Alteio LV, Schulz F, Seshadri R *et al.* Complementary metagenomic approaches improve reconstruction of microbial diversity in a forest soil. *mSystems* 2020;**5**:e00768–19.

Ambler RP. The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 1980;**289**:321–31.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2019. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baldrian P. The known and the unknown in soil microbial ecology. *FEMS Microbiol Ecol* 2019;**95**:fiz005.

BBMap. 2018. https://sourceforge.net/projects/bbmap/.

Behera P, Mohapatra M, Kim JY *et al.* Benthic archaeal community structure and carbon metabolic profiling of heterotrophic microbial communities in brackish sediments. *Sci Total Environ* 2020;**706**:135709.

Biskaborn BK, Smith SL, Noetzli J *et al.* Permafrost is warming at a global scale. *Nat Commun* 2019;**10**:264.

Bonito G, Benucci GMN, Hameed K *et al.* Fungal–bacterial networks in the populus rhizobiome are impacted by soil properties and host genotype. *Front Microbiol* 2019;**10**:481.

Bradley CA, Altizer S. Urbanization and the ecology of wildlife diseases. *Trends Ecol Evol* 2007;**22**:95–102.

Brakhage AA, Thön M, Spröte P *et al.* Aspects on evolution of fungal beta-lactam biosynthesis gene clusters and recruitment of trans-acting factors. *Phytochemistry* 2009;**70**:1801–11.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60.

Burkert A, Douglas TA, Waldrop MP *et al.* Changes in the active, dead, and dormant microbial community structure across a pleistocene permafrost chronosequence. *Appl Environ Microbiol* 2019;**85**:e02646–18.

Cantón R, Akóva M, Carmeli Y *et al.* Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. *Clin Microbiol Infect* 2012;**18**:413–31.

Chaumeil P-A, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;**36**:1925–7.

Chen I-MA, Chu K, Palaniappan K *et al.* The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res* 2021;**49**:D751–63.

Cohen J, Screen JA, Furtado JC *et al.* Recent Arctic amplification and extreme mid-latitude weather. *Nat Geosci* 2014;**7**:627–37.

D'Costa VM, King CE, Kalan L *et al.* Antibiotic resistance is ancient. *Nature* 2011;**477**:457–61.

diCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution. *Microbiol Mol Biol Rev* 2017;**81**:e00019–17.

Erbert M, Rechner S, Müller-Hannemann M. Gerbil: a fast and memory-efficient *k*-mer counter with GPU-support. *Algorithms Mol Biol* 2017;**12**:9.

Evans PN, Parks DH, Chadwick GL *et al.* Methane metabolism in the archaeal phylum bathyarchaeota revealed by genome-centric metagenomics. *Science* 2015;**350**:434–8.

Fedorov A, Konstantinov P. Observations of surface dynamics with thermokarst initiation, Yukechi site, Central Yakutia. In: Phillips M, Springman SM, Arenson LU (eds). *Permafrost, Swets & Zeitlinger*, Lisse, Netherlands, 2003, 240–3.

Finley RL, Collignon P, Larsson DGJ *et al.* The scourge of antibiotic resistance: the important role of the environment. *Clin Infect Dis* 2013;**57**:704–10.

Fuchs M, Nitze I, Strauss J *et al.* Rapid fluvio-thermal erosion of a yedoma permafrost cliff in the Lena River Delta. *Front Earth Sci* 2020;**8**:336.

Fuehrer H-P, Schoener E, Weiler S *et al.* Monitoring of alien mosquitoes in Western Austria (Tyrol, Austria, 2018). *PLoS Negl Trop Dis* 2020;**14**:e0008433.

Galperin MY, Wolf YI, Makarova KS *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 2021;**49**:D274–81.

Gilichinsky D, Wagener S. Microbial life in permafrost: a historical review. *Permafr Periglac Process* 1995;**6**:243–50.

Gilichinsky DA, Vorobyova EA, Erokhina LG *et al.* Long-term preservation of microbial ecosystems in permafrost. *Adv Space Res* 1992;**12**:255–63.

Graham DE, Wallenstein MD, Vishnivetskaya TA *et al.* Microbes in thawing permafrost: the unknown variable in the climate change equation. *ISME J* 2012;**6**:709–12.

Haan TJ, Drown DM. Unearthing antibiotic resistance associated with disturbance-induced permafrost thaw in interior Alaska. *Microorganisms* 2021;**9**:116.

Hatzenpichler R. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol* 2012;**78**: 7501–10.

Hertig E. Distribution of anopheles vectors and potential malaria transmission stability in Europe and the mediterranean area under future climate change. *Parasit Vectors* 2019;**12**:18.

Higgins CS, Avison MB, Jamieson L *et al.* Characterization, cloning and sequence analysis of the inducible ochrobactrum anthropi AmpC β-lactamase. *J Antimicrob Chemother* 2001;**47**:745–54.

Hinsa-Leasure SM, Bhavaraju L, Rodrigues JLM *et al.* Characterization of a bacterial community from a Northeast Siberian seacoast permafrost sample. *FEMS Microbiol Ecol* 2010;**74**:103–13.

Huber I, Potapova K, Ammosova E *et al.* Symposium report: emerging threats for human health: impact of socioeconomic and climate

change on zoonotic diseases in the Republic of Sakha (Yakutia), Russia. *Int J Circumpolar Health* 2020;**79**:1715698.

Huber KJ, Overmann J. Vicinamibacteraceae fam. nov., the first described family within the subdivision 6 acidobacteria. *Int J Syst Evol Microbiol* 2018;**68**:2331–4.

Hueffer K, Drown D, Romanovsky V *et al.* Factors contributing to anthrax outbreaks in the circumpolar north. *Ecohealth* 2020;**17**:174–80.

Ivanushkina NE, Kochkina GA, Ozerskaya SM. Fungi in ancient permafrost sediments of the Arctic and Antarctic regions. In: Castello J, Rogers S (eds). *Life in Ancient Ice*. Princeton, NJ: Princeton Press. 2005, 127–40.

Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* 2006;**72**:1719–28.

Johnson SS, Hebsgaard MB, Christensen TR *et al.* Ancient bacteria show evidence of DNA repair. *Proc Natl Acad Sci USA* 2007;**104**:14401–5.

Jones P, Binns D, Chang H-Y *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40.

Jongejans LL, Liebner S, Knoblauch C *et al.* Greenhouse gas production and lipid biomarker distribution in Yedoma and Alas thermokarst lake sediments in Eastern Siberia. *Global Change Biol* 2021;**27**:2822–39.

Kashuba E, Dmitriev AA, Kamal SM *et al.* Ancient permafrost staphylococci carry antibiotic resistance genes. *Microb Ecol Health Dis* 2017;**28**:1345574.

Keita MB, Hamad I, Bittar F. Looking in apes as a source of human pathogens. *Microb Pathog* 2014;**77**:149–54.

Kochkina G, Ivanushkina N, Ozerskaya S *et al.* Ancient fungi in Antarctic permafrost environments. *FEMS Microbiol Ecol* 2012;**82**:501–9.

Konishi M, Nishi S, Takami H *et al.* Unique substrate specificity of a thermostable glycosyl hydrolase from an uncultured anaerolinea, derived from bacterial mat on a subsurface geothermal water stream. *Biotechnol Lett* 2012;**34**:1887–93.

Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 2018;**46**:e35.

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**:357–9.

Legendre M, Bartoli J, Shmakova L *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci USA* 2014;**111**:4274–9.

Legendre M, Lartigue A, Bertaux L *et al.* In-depth study of mollivirus sibericum, a new 30,000-y-old giant virus infecting acanthamoeba. *Proc Natl Acad Sci USA* 2015;**112**:E5327–35.

Li D, Liu C-M, Luo R *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;**31**:1674–6.

Li W, O'Neill KR, Haft DH *et al.* RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 2021;**49**:D1020–8.

Liu Y, Lillepold K, Semenza JC *et al.* Reviewing estimates of the basic reproduction number for dengue, Zika and chikungunya across global climate zones. *Environ Res* 2020;**182**:109114.

Malavin S, Shmakova L. Isolates from ancient permafrost help to elucidate species boundaries in Acanthamoeba castellanii complex (Amoebozoa: Discosea). *Eur J Protistol* 2020;**73**:125671.

Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 2004;**32**:W327–31.

Mistry J, Chuguransky S, Williams L *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res* 2021;**49**:D412–9.

Morens DM, Daszak P, Markel H *et al.* Pandemic COVID-19 joins history's pandemic legion. *mBio* 2020;**11**:e00812–20.

Mykytczuk NCS, Foote SJ, Omelon CR *et al.* Bacterial growth at −15°C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J* 2013;**7**:1211–26.

Nan F, Jiang J, Wu S *et al.* A novel VIII carboxylesterase with high hydrolytic activity against ampicillin from a soil metagenomic library. *Mol Biotechnol* 2019;**61**:892–904.

Nayfach S, Pollard KS. Toward accurate and quantitative comparative metagenomics. *Cell* 2016;**166**:1103–16.

Nitzbon J, Westermann S, Langer M *et al.* Fast response of cold ice-rich permafrost in northeast siberia to a warming climate. *Nat Commun* 2020;**11**:2201.

Ogawara H. Self-resistance in *Streptomyces*, with special reference to β-lactam antibiotics. *Molecules* 2016;**21**:605.

Oksanen J, Blanchet FG, Friendly M *et al. V e gan: community ecology packag e*. 2022. https://cran.r-project.org.

Pandey SD, Jain D, Kumar N *et al.* MSMEG_2432 of *Mycobacterium smegmatis* mc^2155 is a dual function enzyme that exhibits DD-carboxypeptidase and β-lactamase activities. *Microbiology* 2020:**166**:546–53.

Panikov NS, Sizova MV. Growth kinetics of microorganisms isolated from alaskan soil and permafrost in solid media frozen down to −35°C: growth kinetics in frozen media. *FEMS Microbiol Ecol* 2007;**59**:500–12.

Parkinson AJ, Evengard B, Semenza JC *et al.* Climate change and infectious diseases in the Arctic: establishment of a circumpolar working group. *Int J Circumpolar Health* 2014;**73**:25163.

Pellow D, Mizrahi I, Shamir R. PlasClass improves plasmid sequence classification. *PLoS Comput Biol* 2020;**16**:e1007781.

Perron GG, Whyte L, Turnbaugh PJ *et al.* Functional characterization of bacteria isolated from ancient Arctic soil exposes diverse resistance mechanisms to modern antibiotics. *PLoS One* 2015;**10**:e0069533.

Petrova M, Kurakov A, Shcherbatova N *et al.* Genetic structure and biological properties of the first ancient multiresistance plasmid pKLH80 isolated from a permafrost bacterium. *Microbiology* 2014;**160**:2253–63.

Plowright RK, Parrish CR, McCallum H *et al.* Pathways to zoonotic spillover. *Nat Rev Microbiol* 2017;**15**:502–10.

Pormohammad A, Nasiri MJ, Azimi T. Prevalence of antibiotic resistance in *Escherichia coli* strains simultaneously isolated from humans, animals, food, and the environment: a systematic review and meta-analysis. *Infect Drug Resist* 2019;**12**:1181–97.

Potron A, Poirel L, Nordmann P. Emerging broad-spectrum resistance in *Pseudomonas aeruginosa* and *Acinetobacter baumannii*: mechanisms and epidemiology. *Int J Antimicrob Agents* 2015;**45**:568–85.

R Core Team. 22-04-2022. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org

Revich BA, Podolnaya MA. Thawing of permafrost may disturb historic cattle burial grounds in East Siberia. *Glob Health Action* 2011;**4**:8482.

Revich BA, Tokarevich N, Parkinson AJ. Climate change and zoonotic infections in the Russian Arctic. *Int J Circumpolar Health* 2012;**71**:18792.

Rodriguez-R LM, Gunturu S, Tiedje JM *et al.* Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 2018;**3**:e00039–18.

Roux S, Enault F, Hurwitz BL *et al.* VirSorter: mining viral signal from microbial genomic data. *PeerJ* 2015;**3**:e985.

Sawa T, Kooguchi K, Moriyama K. Molecular diversity of extended-spectrum β-lactamases and carbapenemases, and antimicrobial resistance. *J Intensive Care* 2020;**8**:13.

Schirrmeister L, Froese D, Tumskoy V *et al.* Yedoma: late pleistocene ice-rich syngenetic permafrost of Beringia. In: *Encyclopedia of Quaternary Science*. 2nd edn. 2013, 542–52.

Schneider von Deimling T, Grosse G, Strauss J *et al.* Observation-based modelling of permafrost carbon fluxes with accounting for deep carbon deposits and thermokarst activity. *Biogeosciences* 2015;**12**:3469–88.

Sekiguchi Y, Yamada T, Hanada S *et al.* Anaerolinea thermophila gen. nov., sp. nov. and Caldilinea aerophila gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain bacteria at the subphylum level. *Int J Syst Evol Microbiol* 2003;**53**:1843–51.

Shmakova L, Bondarenko N, Smirnov A. Viable species of flamella (Amoebozoa: Variosea) isolated from ancient Arctic permafrost sediments. *Protist* 2016;**167**:13–30.

Strauss J, Schirrmeister L, Grosse G *et al.* Deep yedoma permafrost: a synthesis of depositional characteristics and carbon vulnerability. *Earth Sci Rev* 2017;**172**:75–86.

Timofeev V, Bahtejeva I, Mironova R *et al.* Insights from *Bacillus anthracis* strains isolated from permafrost in the tundra zone of Russia. *PLoS One* 2019;**14**:e0209140.

Tripathi BM, Kim1 HM, Jung JY *et al.* Distinct taxonomic and functional profiles of the microbiome associated with different soil horizons of a moist tussock tundra in Alaska. *Front Microbiol* 2019;**10**:1442.

Turetsky MR, Abbott BW, Jones MC *et al.* Carbon release through abrupt permafrost thaw. *Nat Geosci* 2020;**13**:138–43.

Turetsky MR, Abbott BW, Jones MC *et al.* Permafrost collapse is accelerating carbon release. *Nature* 2019;**569**:32–4.

Udikovic-Kolic N, Wichmann F, Broderick NA *et al.* Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proc Natl Acad Sci USA* 2014;**111**:15202–7.

Ulrich M, Jongejans LL, Grosse G *et al.* Geochemistry and weathering indices of Yedoma and alas deposits beneath thermokarst lakes in Central Yakutia. *Front Earth Sci* 2021;**9**:704141.

Valentine MJ, Murdock CC, Kelly PJ. Sylvatic cycles of arboviruses in non-human primates. *Parasit Vectors* 2019;**12**:463.

Van Goethem MW, Pierneef R, Bezuidt OKI *et al.* A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* 2018;**6**:40.

Vishnivetskaya TA, Petrova MA, Urbance J *et al.* Bacterial community in ancient siberian permafrost as characterized by culture and culture-independent methods. *Astrobiology* 2006;**6**:400–14.

Vorobyova E, Minkovsky N, Mamukelashvili A *et al.* Micro-organisms and biomarkers in permafrost. In: Paepe R, Melnikov VP, Elfi Van Overloop E *et al.* (eds). *Permafrost Response on Economic Development, Environmental Security and Natural Resources*. Springer: Dordrecht, Netherlands, 2001, 527–41.

Willerslev E, Hansen AJ, Poinar HN. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 2004;**19**:141–7.

Windirsch T, Grosse G, Ulrich M *et al.* Organic carbon characteristics in ice-rich permafrost in alas and Yedoma deposits, Central Yakutia, Siberia. *Biogeosciences* 2020;**17**:3797–814.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;**20**:257.

Wright GD. Environmental and clinical antibiotic resistomes, same only different. *Curr Opin Microbiol* 2019;**51**:57–63.

Xue Y, Jonassen I, Øvreås L *et al.* Bacterial and Archaeal metagenome-assembled genome sequences from Svalbard permafrost. *Microbiol Resour Announc* 2019;**8**:e00516–9.

Yamada T, Sekiguchi Y, Hanada S *et al. Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea ta r divitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *An aer o lineae classis* nov. and *Caldilineae classis* nov. in the bacterial phylum *Chloroflex i*. *Int J Syst Evol Microbiol* 2006;**56**:1331–40.

Yashina S, Gubin S, Maksimovich S *et al.* Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost. *Proc Natl Acad Sci USA* 2012;**109**:4008–13.

Yim G, Wang HH, Davies J. Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci* 2007;**362**:1195–200.

Zhou P, Yang X-L, Wang X-G *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;**579**:270–3.

Zhu W, Lomsadze A, Borodovsky M. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res* 2010;**38**:e132.