High prevalence and diversity of beta-lactamase-encoding bacteria in cryosoils and ancient permafrost

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- Aix Marseille University, CNRS, IGS, Information Génomique & Structurale (UMR7256), Institut de Microbiologie de la Méditerranée (FR 3489), Marseille, 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, Russia
3- Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Permafrost Research Section, 14473 Potsdam, Germany
4- University of Potsdam, Institute of Geosciences, Potsdam, 14478, 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, France

Correspondance to: Prof Jean-Michel.Claverie, Aix Marseille University, Marseille, F-13288, France

E-mail: Jean-Michel.Claverie@univ-amu.fr

Keywords: metagenomic; Kamchatka; Yakutia ; Siberia; permafrost thaw; antibiotic resistance genes

Summary

Background Antimicrobial resistance is one of the major challenges affecting public health. It is mostly due to the continuous emergence of extended-spectrum β-lactamase from various environments followed by their rapid dissemination and selection in clinical settings. The warming of Earth's climate is the other global threat facing human society, in particular with the Arctic regions experiencing a twice faster warming than the global average and permafrost affected by widespread thawing. A potentially dreadful combination of these two threats would be the release and dispersion of harmful microbes that have remained confined to largely uninhabited Arctic regions, or are stored dormant in permafrost.

Methods Environmental DNA was isolated from 12 soil samples from various Arctic and subarctic pristine regions in Siberia (Yakutia and Kamchatka), including 9 permafrost samples collected at various depths. The large datasets obtained from high throughput sequencing was assembled in contigs and their protein-gene contents predicted. We used exhaustive similarity searches to perform taxonomical assignments of bacterial, archaeal, and eukaryotic organisms, as well as DNA viruses. In addition, we specifically identified β-lactamase genes and their prevalence per bacterial genome estimated through the detection of two universal single copy genes.

Findings A total of 9.217 10^{11} bp were exploited, leading to a total of 525,313 contigs at least 5kb in size. The DNA content of the various samples was found to be highly variable, not strictly correlated with the depth or radio-carbon-based deposit age, and most likely linked to the global density of microbes trapped in the corresponding permafrost layers. Bacteria account for more than 90% of the contigs in most samples, followed by Eukaryotes and Archaea (always lower than 10%). Viruses represented less than 2% of all contigs in all samples. The taxonomic profiles of surface cryosoils and deep permafrost samples exhibited a high diversity, including between permafrost samples originating from various depths in the same borehole. In all samples, bacterial contigs carrying different β-lactamases from class A to D were identified.

Interpretation No clear common taxonomic feature could be found shared by surface cryosoils or ancient permafrost layers. However, most samples (9/12) exhibited a high frequency of β-lactamase genes, with an estimated average close to 1 copy/bacterial genome. In addition to the well-documented reactivation of infectious ancient pathogens (bacteria, viruses, protozoa), we show now that global warming could contribute to the emergence of new antibiotic resistances through the mobilization by contemporary bacteria of ancient DNA released from thawing permafrost.

Funding CNRS PRC research grant (PRC1484-2018) to C.A. E C-F was supported by a PhD grant (DGA/DS/MRIS) #2017 60 0004. GG and JS were funded by ERC PETA-CARB (#338335) and the HGF Impulse and Networking Fund (ERC-0013).

Introduction

Global warming is on the front stage of societal concerns world-wide together with the increasing frequency of emerging infectious diseases turning into pandemics (such as the latest of Covid-19)(1,2). It turns out than both predicaments are partly linked. Through their influence on the ecology of living organisms transmitting infectious pathogens, a warming climate strongly contributes to the spread of vector-borne diseases endemic to tropical regions to more temperate regions of the globe (3-6). This phenomenon is further amplified by the encroachment of pristine areas (e.g tropical forests) by ever expanding human activities generating new risks through imponderable contacts with (mostly uncharted) microbial environments and their associated wild life hosts (7-10).

While such danger is mostly pointed out (particularly in Western Europe) 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 (11-14).

Climate warming is particularly noticeable in the Arctic where average temperatures are increasing twice as fast as in temperate regions (15). One of the most visible consequences is the widespread thawing of permafrost at increasing depths (16, 17) and the rapid erosion of permafrost bluffs (e.g., 18), a phenomenon most visible in Siberia where deep continuous permafrost underlays most of the North Eastern territories.

The “ice age bug” concern is periodically brought back to the public attention, when exceptionally hot summers trigger local outbreaks of anthrax, such as in 2016, when 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 of Yamal Peninsula, Northwest Siberia, since 75 years (19). Historically frequent 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 settlement across the Russian North (20). Yet, anthrax is a non-contagious bacterial disease curable today given proper medical attention (antibiotics and antitoxin) and for which a vaccine is available.

Of greater concern would be the long-term preservation of animal/human viruses in permafrost. Such possibility became more realistic after the detection of variola virus DNA in a 300-year-old frozen mummy from a burial site in the Churapchinsky District of Yakutia (20, 21). However, no virus particle was likely to be infectious anymore as the viral DNA was found to be fragmented into pieces less than 2 kb (22). Such incomplete preservation could be expected given the shallow depth of the grave, at the boundary of the permafrost active layer (i.e. warming up and thawing every summers).

The capacity of deeper permafrost (hence much older) layers to preserve the integrity of “live” (i.e. infectious) viral particles, was finally demonstrated by the isolation of two previously unknown giant DNA viruses: Pithovirus (23) and Mollivirus (24). Those

Acanthamoeba-infecting viruses were isolated (and cultivated) from a wall-sampled layer of a well-studied permafrost bluff (the Stanchikovskiy Yar) (Sample B, Table 1) radiocarbon dated to 35,000 years BP (23). This first demonstration that eukaryotic DNA viruses could still be infectious after staying dormant since the late Pleistocene came after many other studies (25-27) going back to 1911 (reviewed in 28). It was shown that ancient bacteria could be revived from even much older permafrost layers, up to 2 million years (29), although contaminations by modern bacteria have been suspected (30). A flower plant was also revived from 30,000 year old frozen tissues (31), as well as various fungi (from Antarctica) (32) and amoebal protozoans (33, 34).

There is now little contest 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 these viruses and bacteria 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, carbon dioxide) upon thawing? In order to answer these questions, a first step is to document what is frozen in there. Starting this, the technique of choice (and safest) for initiating this task is metagenomics. In contrast with targeted approaches, such as metabarcoding (i.e. PCR-amplified subset of environmental DNA), the bulk sequencing of non-amplified environmental DNA allows the detection of any living organisms (and virus) without restricting the analysis to a given domain or phylum. It thus preserves the possibility of unexpected discoveries. Without amplification, it also provides an image of the real abundance of the various species, although at the cost of a lower sensitivity. Finally, the analysis of metagenomic sequences obtained from gram quantities of ancient permafrost samples is much less sensitive to eventual contaminations and more suitable for site-to-site comparisons than the use of culture-based approaches.

Here we report the metagenomics analysis of 9 sub-surface Siberian samples, all taken from carbon-rich, ice-rich frozen soils called Yedoma deposits (35, 36) found in vast regions of northeast Siberia (Lena and Kolyma river basins in Yakutia)(Fig. S1). Yedoma permafrost may be especially prone to rapid thaw processes such as thermokarst and thermo-erosion (37) releasing not only carbon as greenhouse gases (38, 39) but also its formerly freeze-locked microbial content, in which we detected an unexpected high abundance of β-lactamases genes carried in the genomes (or plasmids) of a large phyletic diversity of cryosoil bacteria.

Methods

Sample collection

We sampled Yedoma and thermokarst samples from Central Yakutia from the Yukechi alas landscape about 50 km southeast of the capital of Yakutia, Yakutsk (Fig. S1). Yukechi is characterized by Yedoma uplands and drained as well as recent lake basins, indicating active permafrost degradation processes (40). Field work took place in March 2015 during a joint

Russian-German drilling expedition. Four 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 on the Yedoma upland lake (61.76086°N, 130.47466°E), and one from the adjacent drained Yukechi alas basin (61.76490°N, 130.46503°E). Biogeochemical characterization as well as greenhouse gas production under ponding water are published in (41, 42). The Stanchikovskiy Yar sample (Sample B) from which two live viruses were previously isolated (23, 24), and three modern surface soils from pristine cold regions (Kamchatka), were also analyzed for comparison (Table 1). Fig. S1 indicates the approximate locations of the sample collections.

DNA extraction

Samples were processed using either the miniprep (0.25 g of starting material) or maxiprep (20 g of starting material) version of the the Dneasy PowerSoil Kit from Qiagen. Samples B, C, D, E were processed using the miniprep kit, following the manufacturer protocol and by adding DTT to the C1 buffer at a 10 mM final concentration. Extractions were repeated until 1 µg of raw DNA was obtained for each sample. 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 10mM final. Samples were grounded for 20 seconds in an MP FastPrep homogenizer at a speed of 4 m/s then incubated for 30 minutes at 65 ° C and finally grounded again for 20 seconds 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 NEB.

DNA Sequencing, assembly, and annotation

Each metagenomic sample was sequenced with DNA-seq paired-end protocol on Illumina HiSeq platform at the French National Sequencing Center "Genoscope" (https://jacob.cea.fr) producing datasets of 2x150bp read length, except for sample B (2x100bp read length). Reads were quality checked with FASTQC (v0.11.7) (43). Identified contaminants were removed and remaining reads were trimmed on the 3' end using 30 as quality threshold with BBTools (v38.07) (44). Assemblies of filtered datasets were performed using MEGAHIT (v1.1.3) (45) 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 (46) with the --very-sensitive option. To ensure a reliable taxonomic assignation, only scaffolds longer than 5kb were considered and submitted to metaGeneMark (v3.38) (47) with the default metagenome-style model. The predicted genes were then compared to the Refseq database (48) using the Diamond version of blastp (v0.9.31) (49) with an E-value threshold of $1e^{-5}$ and retaining only the best hit. The taxonomy of each scaffold was then inferred with a custom-made script applying a LCA like method. At each rank, recursively, the taxonomy was conserved only if at least half of the annotated genes presented the same keyword. Unclassified scaffolds and potential chimeras were screened against Refseq with Diamond blastx (--sensitive option and E-value of $1e^{-5}$), 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) (50) using its standard database. This annotation was found to be much less comprehensive and less reliable and was not used further.

β-lactamase gene detection and analyses

ORFs with a best Diamond blastp match to a β-lactamase in the NCBI Refseq database (see above) were retrieved and analysed. Functional domains were confirmed/inferred by Interproscan (v5) (51) against Pfam, ProSitePatterns, and by Batch-CDsearch (52) against the COG database (53) (E-value < 10^{-5}). 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 (54) and PlasFlow (55). However, the results were found to be unreliable and not used further. To corroborate the high proportion of β-lactamases genes initially computed from contigs larger than 5 kb in size, the same protocol was reiterated on an independent dataset constituted of the smaller contigs in the 2kb-5kb size range. It was further confirmed by analysing a smaller open-access metagenomic dataset obtained from Arctic permafrost samples from three different locations.
Within each dataset**,** β-lactamase counts were normalized against three independent estimates of the number of equivalent complete bacterial genomes, using the detected number of two single copy marker genes : Phenylalanine—tRNA ligase (Uniprot accession W9B9D0) and 50S ribosomal protein L2 (NCBI accession WP_000301869.), as well as mean size of bacterial genomes. The marker proteins were first aligned to our datasets by blastP (E-value <1e-5), the resulting matching ORFs then confirmed by comparison to Refseq, and counted.

The sequences of the β-lactamase-encoding contigs identified in this work will be made freely available upon publication.

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	-	35,000	Wall sampling, Yedoma bluff
Yuk15-Yed-1 /Yed-6	K	61.75967	130.47438	Yukechi Alas landscape	16	>49,000	Core, Yedoma upland
Yuk15-Yed-1 /Yed-7	P	61.75967	130.47438	Yukechi Alas landscape	11	>36,000	Core, Yedoma upland
Yuk15-Alas-1 / Ala-11	L	61.76490	130.46503	Yukechi Alas landscape	12	>28,000	Core, from a drained thermokarst lake basin (Alas)
Yuk15-Alas-1 / Ala-10	N	61.76490	130.46503	Yukechi Alas landscape	16	>45,000	Core, from a drained thermokarst lake basin (Alas)
Yuk15-Alas-1 / Ala-9	R	61.76490	130.46503	Yukechi Alas landscape	19	>42,000	Core, from a drained thermokarst lake basin (Alas)
Yuk15-YuL-15/Y1	O	61.76086	130.47466	Yukechi Alas landscape	19	>40,000	Core, Yedoma upland, under lake (4.5 m water depth)
Yuk15-YuL-15/Y2	M	61.76086	130.47466	Yukechi Alas landscape	16	>48,500	Core, Yedoma upland, under lake (4.5 m water depth)
Yuk15-YuL-15/Y4	Q	61.76086	130.47466	Yukechi Alas landscape	6	53[#]	Core, Yedoma upland, under lake (4.5 m water depth)

1: this soil sample was not frozen (talik). Datat of O, M, Q: 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 (41, 42).

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

Table 2: Raw sequencing data

Sample code	Source location	Dating (years BP)	µg DNA/ g sample	Total bp (250 ng)	Total reads (250 ng)	N contigs (>5kb)	Average size (>5kb)	Median size	Median coverage
C	Kamchatka	present	10.24	82,686,747,816	676,965,994	42,130	10,314	7205	10.7
D	Kamchatka	present	0.93	85,127,467,969	642,216,312	78,167	11,551	7665	12.7
E	Kamchatka	present	1.91	86,068,819,231	685,386,422	16,641	8,659	6638	12.9
B*	Stanchikovskiy Yar	35,000	4.5	16,485,334,216	222,759,958	15,293	15,188	8583	13.3
L	Yukechi Alas	>28,000	0,023	83,976,217,055	644,805,720	34,431	12,054	7717	20.1
R	Yukechi Alas	>42,000	0.054	81,053,849,744	627,765,894	35,616	12,237	7787	19.9
N	Yukechi Alas	>45,000	0.020	89,953,345,271	674,592,444	28,197	11,450	7782	28.1
P	Yukechi Yedoma	>36,000	0.060	80,623,042,762	615,573,200	64,903	9,449	7168	16.3
K	Yukechi Yedoma	>49,000	0.065	71,912,512,659	569,204,104	49,767	10,051	7172	12.9
Q	Yukechi Yedoma lake	53	0.067	82,881,074,132	635,044,766	38,175	15,803	8256	19.9
M	Yukechi Yedoma lake	>48,500	0.049	83,875,504,131	648,500,972	58,035	9,764	7248	15.4
O	Yukechi Yedoma lake	>40,000	0.051	77,041,387,804	585,962,630	63,958	9,802	7313	12.9

* 100 bp/reads

Results

DNA content

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 1, for instance sample B *vs*. P). A lesser range of variation (10 times) was observed between surface samples (Sample C to 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 2). 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 2), indicating that, once purified, all DNAs exhibited a similar quality. 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.

In contrast, the assembly of the similar number of sequence 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 is a first indication that the samples exhibit globally different microbial population compositions and structures. This was readily confirmed by the histogram of the G+C composition of the various contigs (Fig. 1).

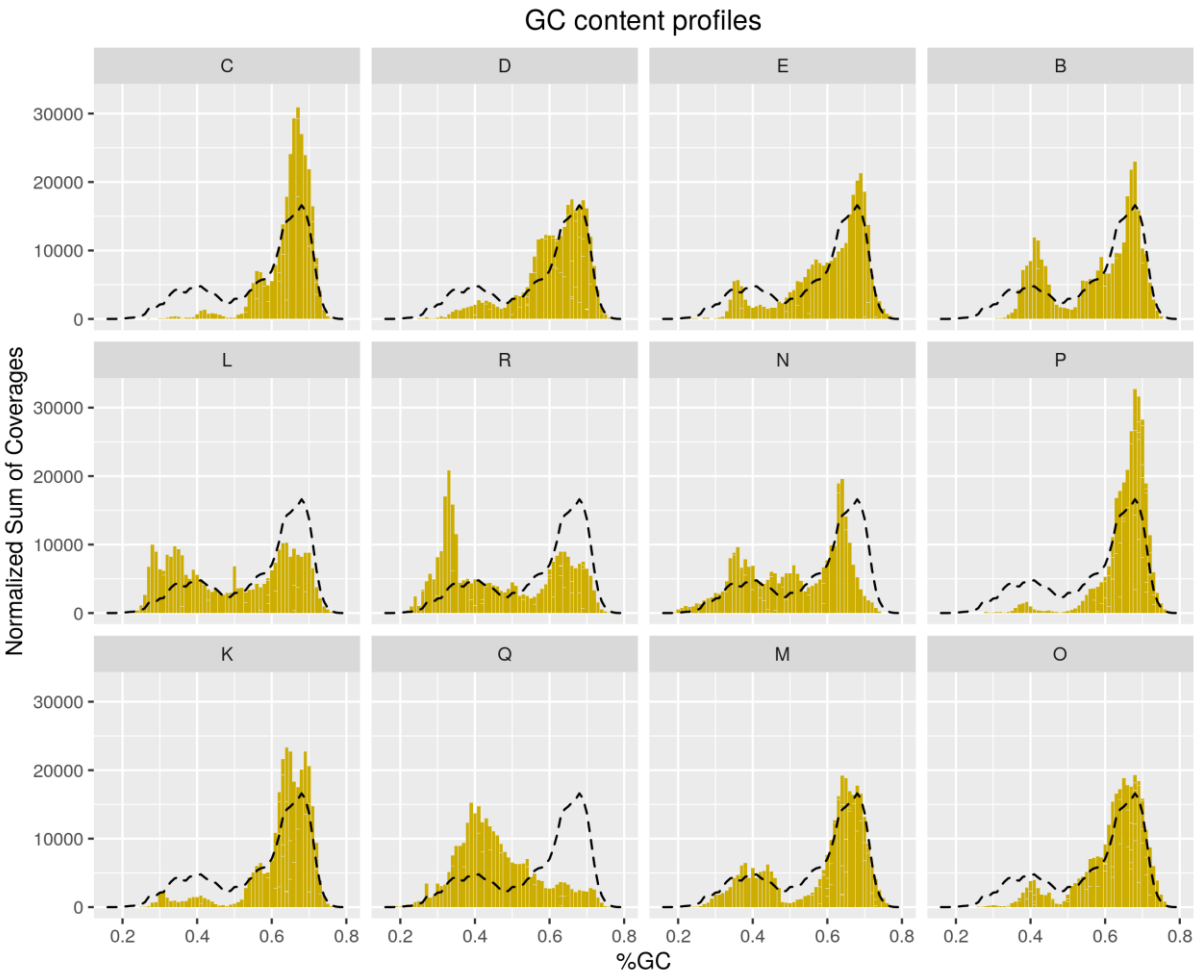


Figure 1

***Legend**: Global G+C content (%) in each sample, computed from their total usable bp contents. The profiles are markedly different, including for samples extracted from various levels in the same borehole: L,N,R; O,Q; K,P;. This is consistent with a minimal mixing of the microbial content. The dashed line indicates the average GC profile all samples combined.*

Figure 1 illustrates the abundance of the various organisms associated with a given G+C% value in the various sample. Most of them exhibit a bimodal distribution with two peaks centered around 40% and 67% G+C, although with large variation of their relative proportions. Some of the sample exhibit almost no A+T rich organisms (e.g. C and P), while the predominant peak of G+C rich organism appears much reduced in others (e.g. L, Q, R). In two extreme cases (e.g. P and Q), a single of these peak was present suggesting the absence of whole classes of bacteria (Fig. 1). Adjacent samples from the same borehole exhibit different profiles (e.g. L, N, R and Q) supporting the absence of large-scale homogenization of the microbial populations between strata of ancient frozen soils. This validates the notion that the DNA detected at various depths truly relates to the microbes syngenetically trapped at the time of permafrost formation. The large variations of the G+C profiles between samples suggest - prior to any taxonomical annotation -, that ancient permafrost layers could harbor vastly different microbial populations. Upon thawing, the ecological or medical consequences of their eventual reactivation and release probably do not lend themselves to generalization.

Global species contents

Compared to bar-coding methods, metagenomics allows an exploration of the sample diversity that is, in principle, only limited by a threshold of minimal natural abundance for a given species. This threshold is in fact a composite of several factors mixing the physical proportion of cells (or viral particles) of a given species (mostly unicellular) in each sample with other parameters strongly influencing our capacity to identify them by sequence similarities searches in the genomic database of known organisms. These parameters include the protein-coding density (high in viruses and prokaryotes, much lower in most eukaryotes), the fraction of the database corresponding to known members of the various phyla (high for bacteria, much lower for viruses, archaea and eukaryotes), the level of sequence similarity among homologous genes in the different domains (low in viruses, higher but variables in cellular organisms), and the proportion of ORFans in the various domains (low in Eukaryotes, high in prokaryotes, highest in viruses). Thus, for a similar coverage (i.e. relative proportion) in a metagenomics mixture, bacteria are more likely to be identified than Archaea and Eukaryotes, with viruses being the most likely to be overlooked (and remaining taxonomically unclassified). The comparisons of organism frequencies across domains are thus less reliable than within domains. Yet, the overwhelming predominance of DNA associated to the bacterial domain that characterizes all the samples is unlikely to be due to a bioinformatic/annotation bias. As depicted in Figure 2, bacteria account for at least 95% of contigs (> 5kb) in most samples. This is true of all modern cryosoils and most ancient permafrost layers despite their different origins and DNA contents. A similar predominance for bacterial was indicated using a popular read-based annotation method (Fig. S2) (50). However, we found this method to be less sensitive in identifying other cellular domains and viruses and was not used further in this study.

Eukaryotic DNA is representing less than 1% of contigs, except for two sub-surface samples from the same borehole (R, N). The predicted taxonomy of these eukaryotic organisms will be described in a specific section.

A group of three samples (L, R, N) collected from the same borehole at three different depths (respectively 12 m, 16 m, 19 m) stand out from the rest by their higher proportions of contigs of archaeal origins or annotated as "chimera". These three samples correspond to the largest median coverage values, suggesting a lower diversity of organisms (i.e. smaller richness). Contigs predicted as "chimera" probably originate from various phages the detection of which is known to be difficult in metagenomics data. A specific section will describe the types of archaea detected in the above samples.

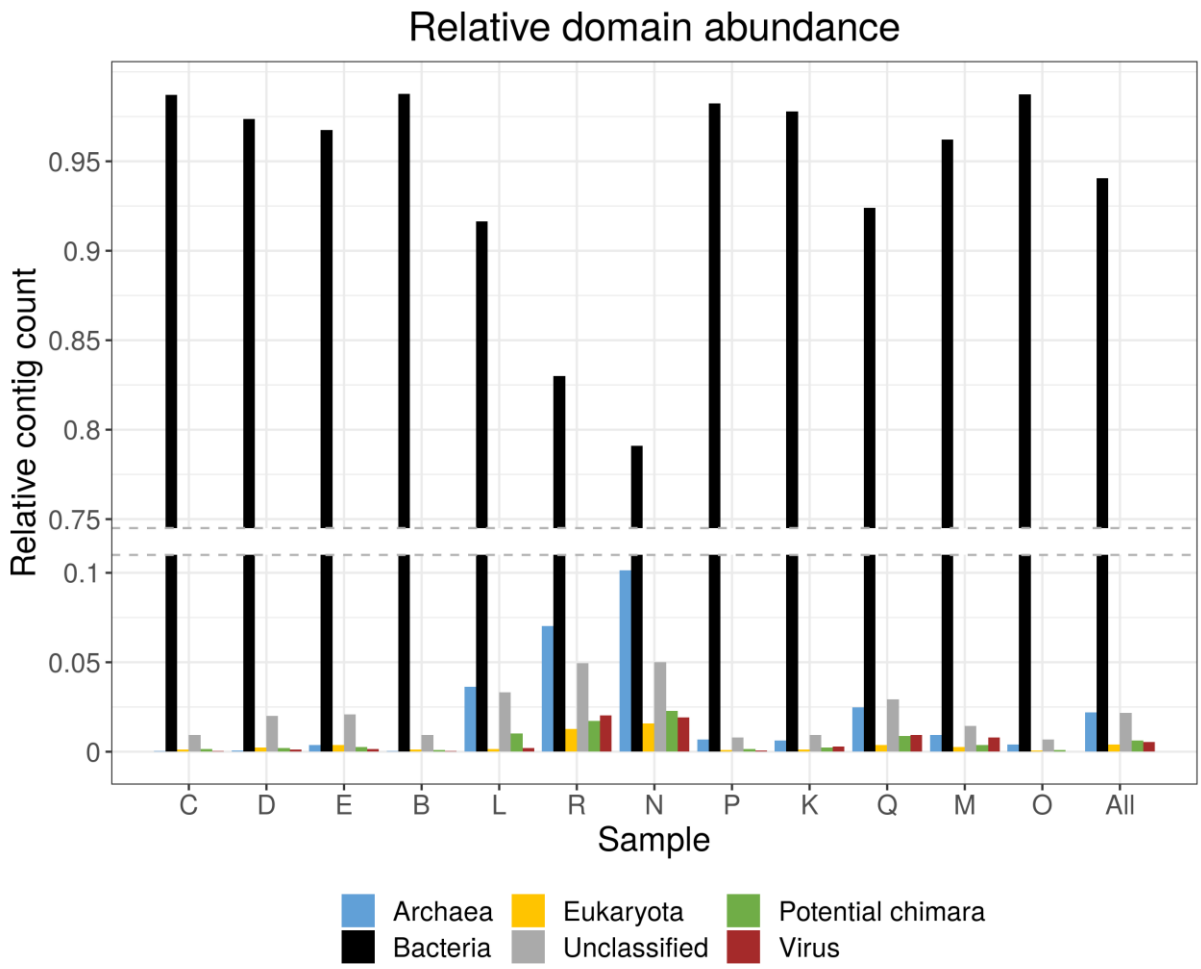


Figure 2.

***Legend**: Relative distributions of organism types identified in the various samples. Notice that the absolute number of contigs largely varies between samples (Table 2).*

Focus on the bacterial populations

The relative abundances of the phyla representing more than 0.5% of the contigs are depicted in Fig.3.

Even at this high taxonomical level, the picture is quite variable from one sample to the next, making it illusory to expect that a tractable number of bacterial genera could be used to classify the various types of cryosoil and permafrost into recurrent taxonomical or functional types. For instance, bacterial populations from superficial cryosoils (C, D, E) are not more similar to each other than they appear to be in ancient permafrost. Similarly, samples originating from different depths in the same borehole (such as L, R, N) do not exhibit strong similarity (ruling out by the same token the possibility of cross-contamination). Although Proteobacteria appear globally dominant (40% of the total contigs), it is not always the most abundant in every sample. The most extremes exceptions 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 also vary strongly across samples from up to 24%, 18%, 13% and 8.7%, (respectively) down to almost 0. None of these characteristics appear to correlate with the geographic location or age of the ancient samples.

Even within a given phylum, such as the Proteobacteria (Fig. 4), very distinct distributions across classes could be seen between different modern cryosoils (C, D, E) or ancient permafrost layers either from the same borehole (L-R-N; Q-M-O), or from similar depth (19m: 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.

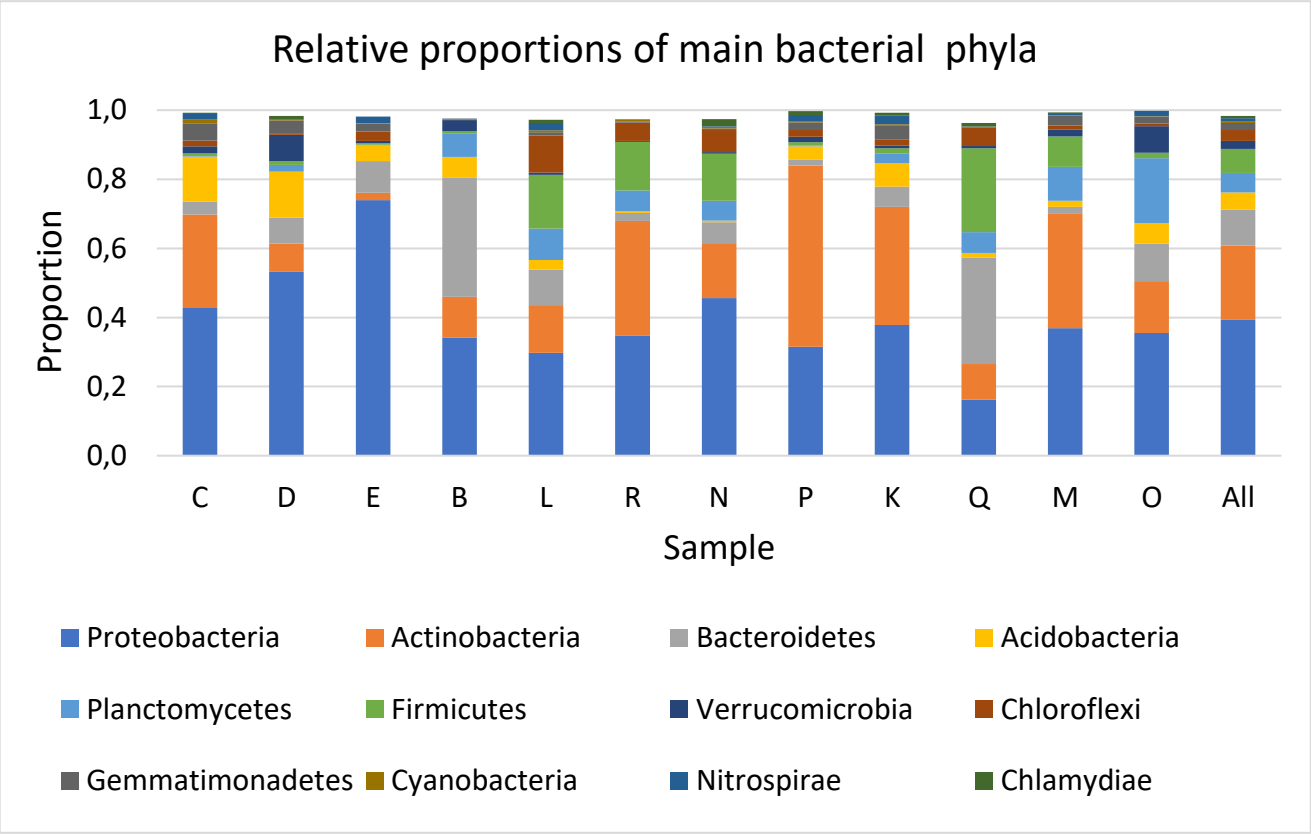


Figure 3.

***Legend**: Relative distributions of bacterial phyla in the various sample. Only phyla representing more than 0.5% of bacterial contigs are indicated.*

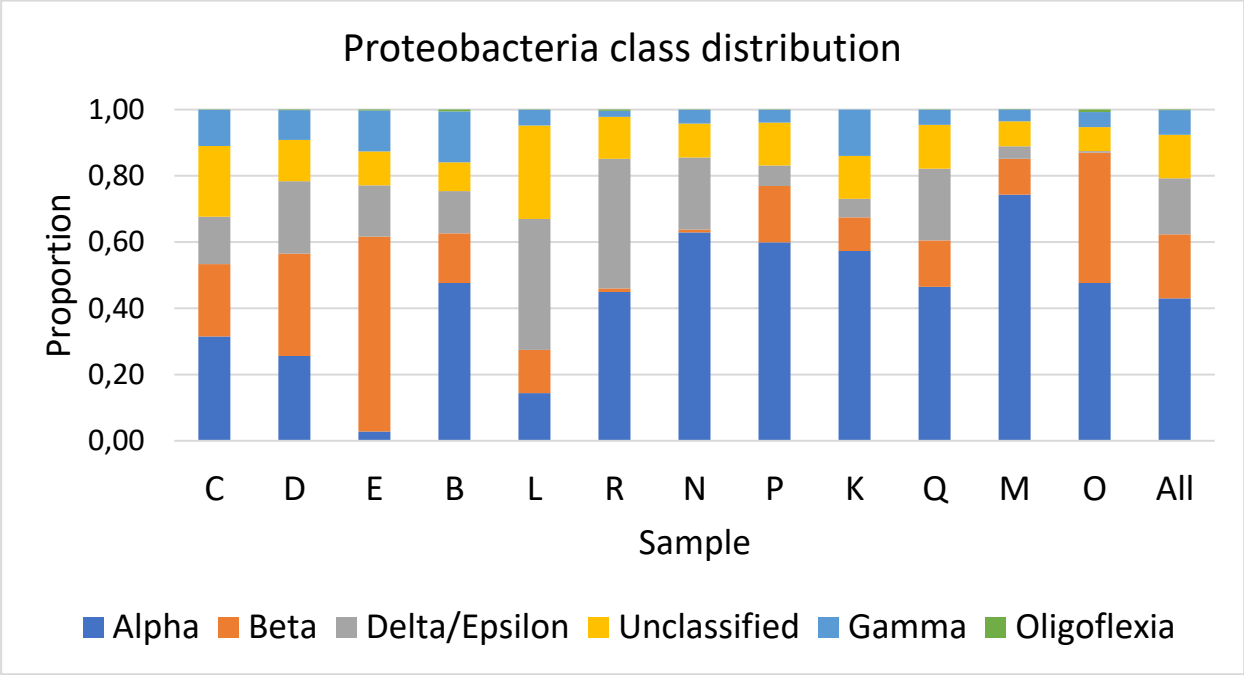


Figure 4.

***Legend**: Relative distributions of the main clades within the dominant Proteobacteria phylum. Samples originating from various layers of the same borehole may exhibit very different profiles (e.g. L,R,N; Q,M,O), demonstrating the absence of significant mixing of their microbiome.*

Focus on the archaeal populations

The relative abundance of the main classes, orders, or phylum (at least 0.6% of the archaeal contigs) in the various samples are depicted in Fig. 5.

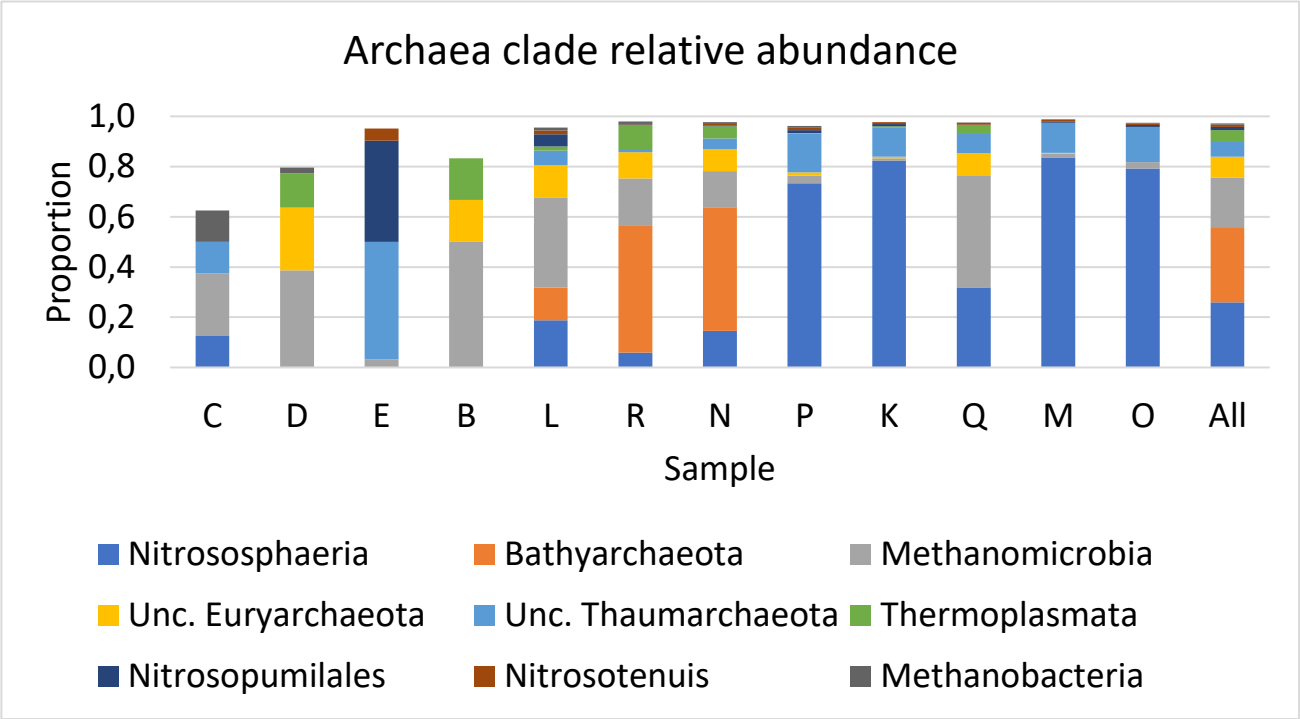


Figure 5

Legend*: Relative distribution of the main clades of the Archaeal domain. The distributions are even more diverse than those computed from bacteria abundances. However, it must be noticed that the absolute number of archaeal contigs is much smaller, in particular for sample C, D, E, B (respectively 16, 44, 62 and 6 total archaeal contigs, Fig. 2).*

This bar graph should be interpreted while keeping in mind that Archaeal contigs constitute less than 1 in thousands of all contigs in most samples (Fig. 2), except for samples L (3.5%), N (9%), R (6.75%) and Q (2.25%). As expected from anaerobic microbes, archaea are virtually absent from surface samples (less than 6/10.000 for C, D, and 4/1000 for E). In the sample where they are in sizable amounts, the archaeal population is dominated by Bathyarcheota (a recently described clade of methanogens belonging to the TACK phylum) (56), Nitrososphaeria (a class of chemolithoautotrophic ammonia oxidizing archaea belonging to Thaumarchaea phylum and exhibiting global distribution in soils) (57), and Methanomicrobia (a class of methanogens belonging to the Euryarchaeota and previously noticed to be enriched in macrophytes rhizosphere sediments) (58).

Focus on the Eukaryote populations

The relative abundance of the main clades of Eukaryotic organisms (in term of relative number of contigs larger than 5 kb) is shown in Fig. 6. As previously (for Archaea), this bar graph should be interpreted keeping in mind that Eukaryotic contigs represent less than 1% of all contigs in most samples (Fig. 2), except for samples R (1,7%) and N (2.1%).

Despite a large variation in the proportion of Eukaryotic contigs between the different samples, the viridiplantae (land plants and green algae) and unikont (mostly fungi, metazoan and amoebozoa) constitute most of the represented clades except for sample E and O. If the later 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 the 40,000 year-old layer corresponding to this sample was located at the bottom of a lake (Table 1A).

Concerning contigs associated to unikont organisms, the three main categories are metazoans, fungi, and amoebozoa (Fig. 7). Their relative are quite variable across different samples (although some are computed from a very small number of contigs: 12 for B and O, see Fig. 2).

Among the samples exhibiting more than 100 unikont contigs (D, R, N, Q, M) the distributions remain quite variable, in particular between two superficial samples, both originating from Kamchatka (C, E).

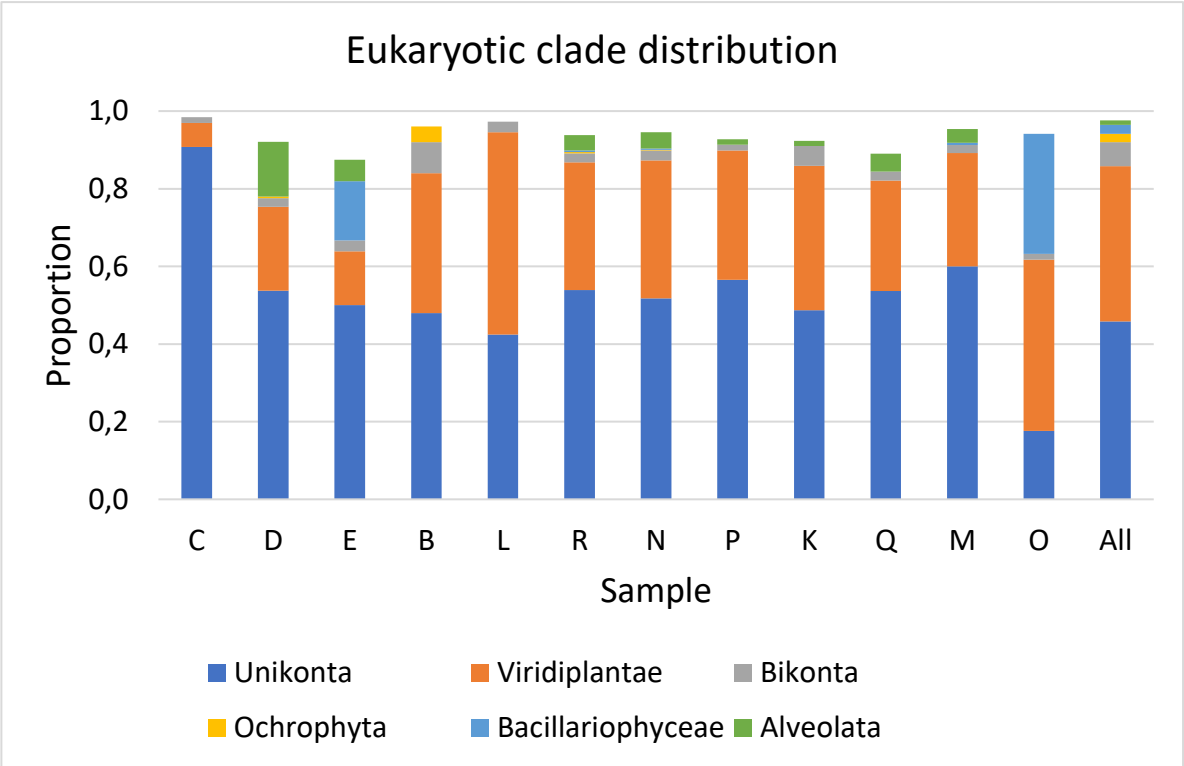


Figure 6

***Legend**: Relative distribution of the main clades of the Eukarya domain. It must be noticed that the absolute numbers of eukarya contigs is very small in most samples (less than 1% except for sample R and N, Fig. 2).*

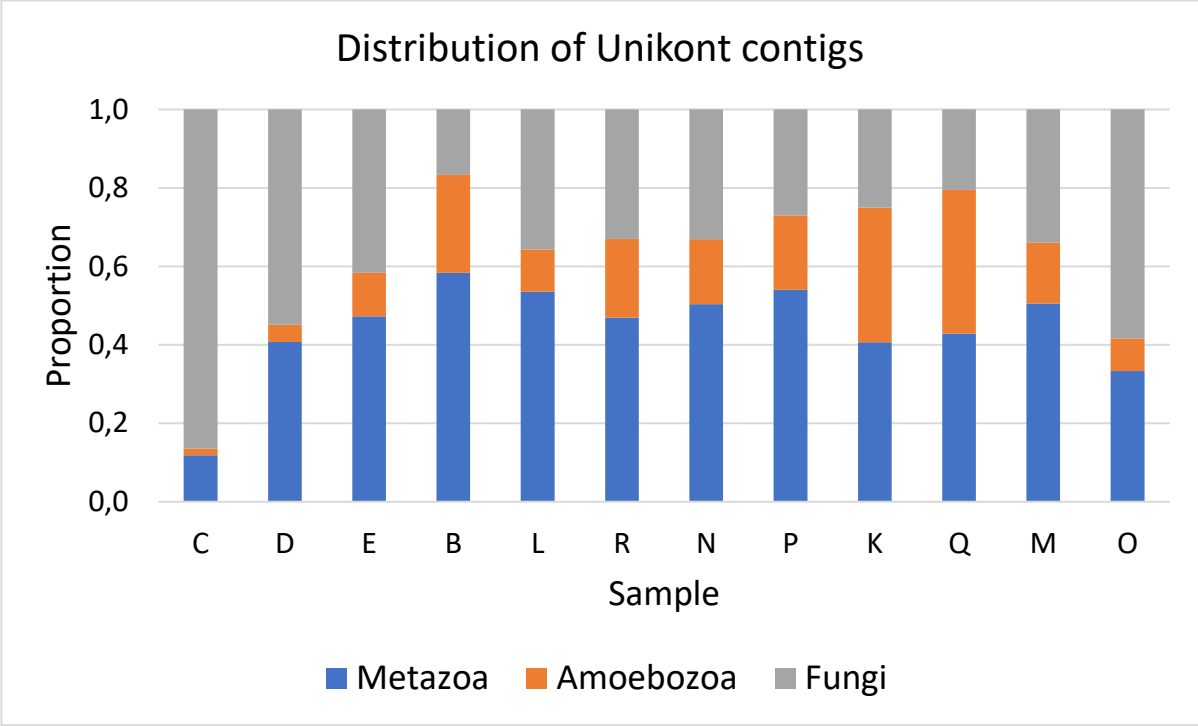


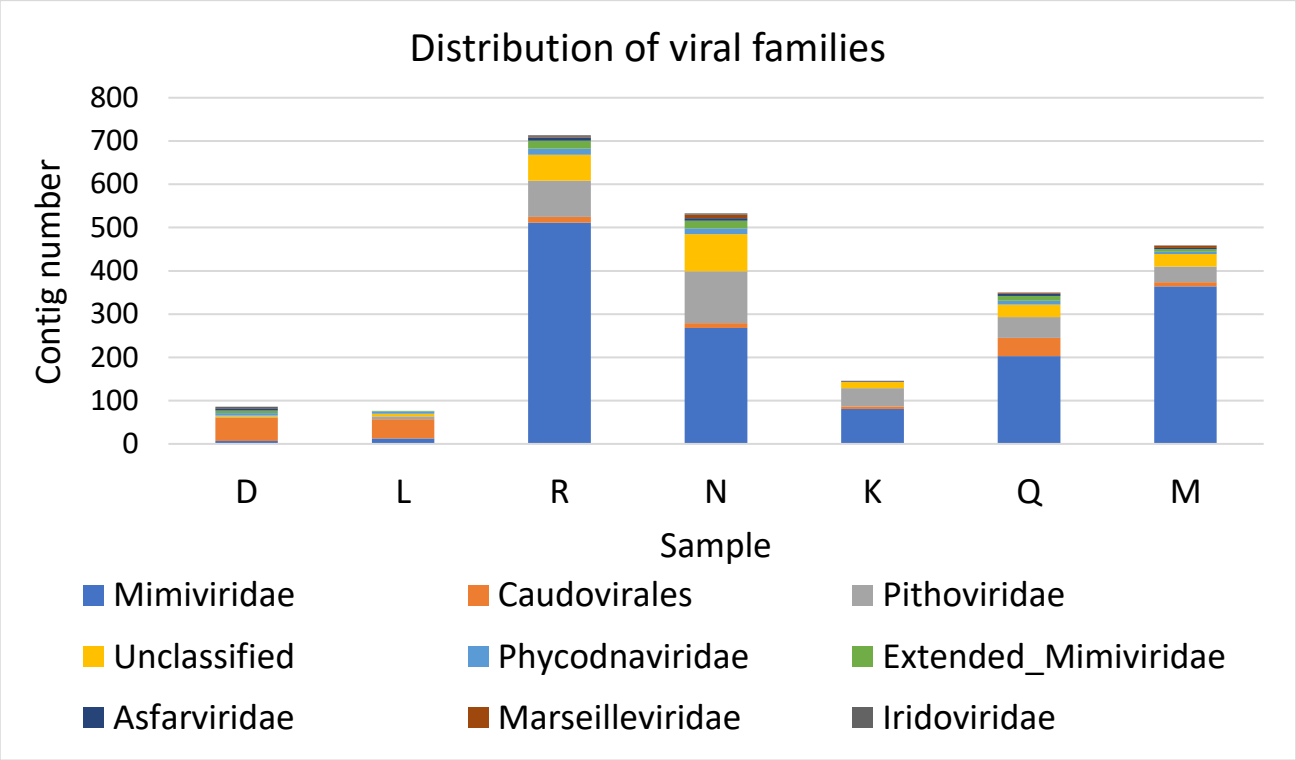
Figure 7

***Legend**: Relative distributions of the main clades within the Unikont phylum.*

Focus on the viral populations

At a time when pandemics caused by “new” viruses seems to occur at an accelerated pace, the analysis of the viral populations present in permafrost which is known to be impacted by enhanced and deep thaw and therefore threatens to release them, is of particular interest. In particular, deep thaw by rapid thermokarst or thermo-erosion processes allows remobilization of ancient viral material tens of millennia old within short annual to decadal time scales. Unfortunately, our experimental approach is only opening a window on viruses the genomes of which consist of dsDNA, while the most human threatening viruses are RNA viruses (e.g. Coronaviruses, Influenza viruses, Ebolaviruses). These viruses are also thought to be less resistant to long period of congelation in the environment, although there is no published data on the subject to our knowledge despite the availability of potential research materials (59). We previously demonstrated the long term survival of two different types of giant DNA viruses from a sample of 28,000 year-old permafrost (23, 24) and were eager to explore how general and how diverse was the presence of DNA viruses at various locations. To our surprise, most of the samples we analyzed revealed only trace amount of viral DNA (Fig.2). Contigs classified as viral represented 0.29% of the contigs of all samples while viruses (e.g. bacteriophages) are reputed to be the most numerous microbes in soils (reviewed in 60). Viruses represented close to or more than 1% of all contigs in only 4 samples: R (2%), N (1.9%), Q (0.9%), and M (0.8%). Fig. 8 shows the proportion of main families of viruses (with more than 10 total occurrences across all samples) in the samples for which at least 40 viral contigs were detected. Although the family distributions between theses samples were quite variable, the one with the largest number of viral contigs exhibited a clear predominance of Mimiviridae, followed by Pithoviridae (the prototype of which was previously isolated from ancient permafrost (23), unclassified viruses, and a smaller proportion of phages (Caudovirales).

Figure 8



***Legend**: Relative distributions of the main DNA virus families. It must be noticed that viral contigs represent less than 1% of all contigs in most samples.*

We found these viruses absent from most surface samples. Thus, we suggest that the burying of the cryosoil and its transformation into permafrost is accompanied by viral proliferations, whose genomic traces remain detectable after tens of thousands of years. As previously shown, some of these viruses can still be infectious (23, 24)). The predominance of viruses of the Mimiviridae and Pithoviridae families in ancient permafrost layers correlates with the presence of their most usual cellular hosts (amoebal protozoans) in the same samples (Fig. 7). Live amoeba are routinely isolated from permafrost (33, 34). Except for a small number of contigs associated to Asfarviridae and even less to Iridoviridae, almost no eukaryotic viruses were detected from the families known to infect metazoan organisms, although their metazoan DNA was readily detected in most samples (Fig. 7).

An unexpected large proportion of antibiotic resistance genes

Beyond the direct infectious risk represented by the re-circulation of still viable ancient microorganisms released upon thawing of old 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 taken into account. Although it originated from dead (or even disappeared) organism 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. Of particular interest are DNA fragments coding for bacterial virulence factors, and first of all, antibiotic resistance genes the origin of which has been shown to predate by far human history (61-64). Moreover, we focused our analyses on β-lactamases the specific sequence motifs of which are readily identifiable, and for the central role the continuous emergence of extended-spectrum β-lactamases (ESBL) (64-66) play in the spread of bacterial antibiotic multiple resistances.

Fig. 9 presents the number of bacterial contigs within which β-lactamases of the 4 classes (according to Ambler's classification) (67) were detected with high confidence (Fig. S3). Although these absolute numbers appear small compared to the total number of bacterial contigs (Table 2 and Fig. 2), they suggest that most of the sampled bacteria harbours close to one copy of a β-lactamase gene, once normalized by the frequency of two known single copy number genes detected in each sample data set (Fig. 10).

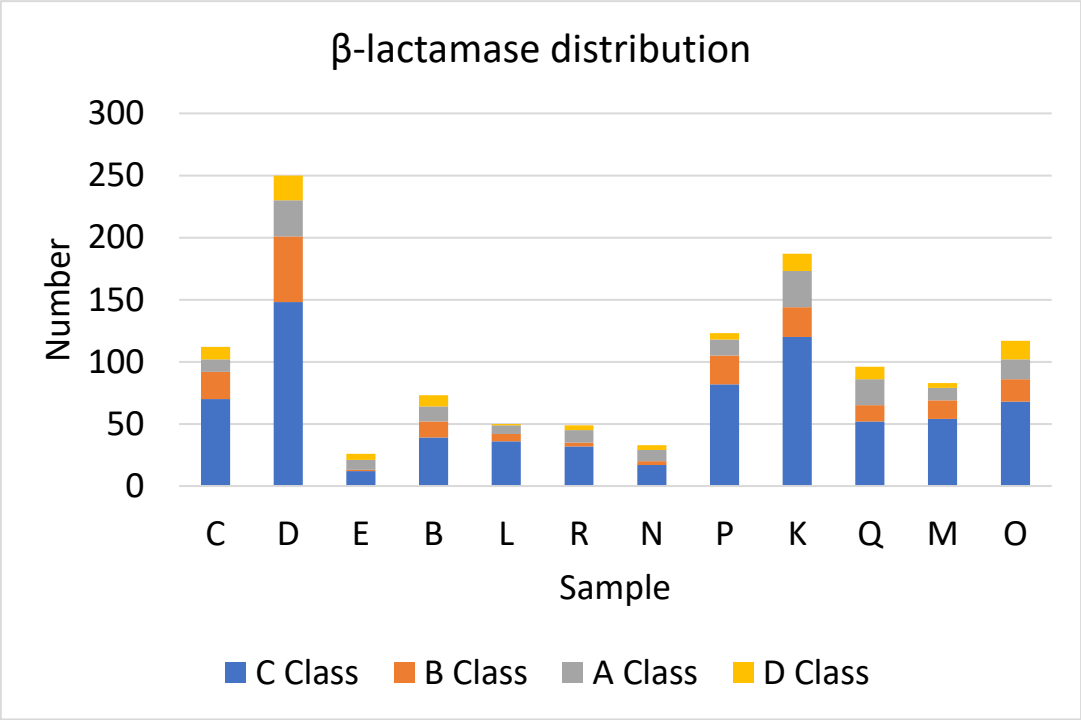


Figure 9

***Legend**: Relative distributions of the 4 main Ambler's classes of β-lactamases in the various samples. See detailed statistics in Table S1.*

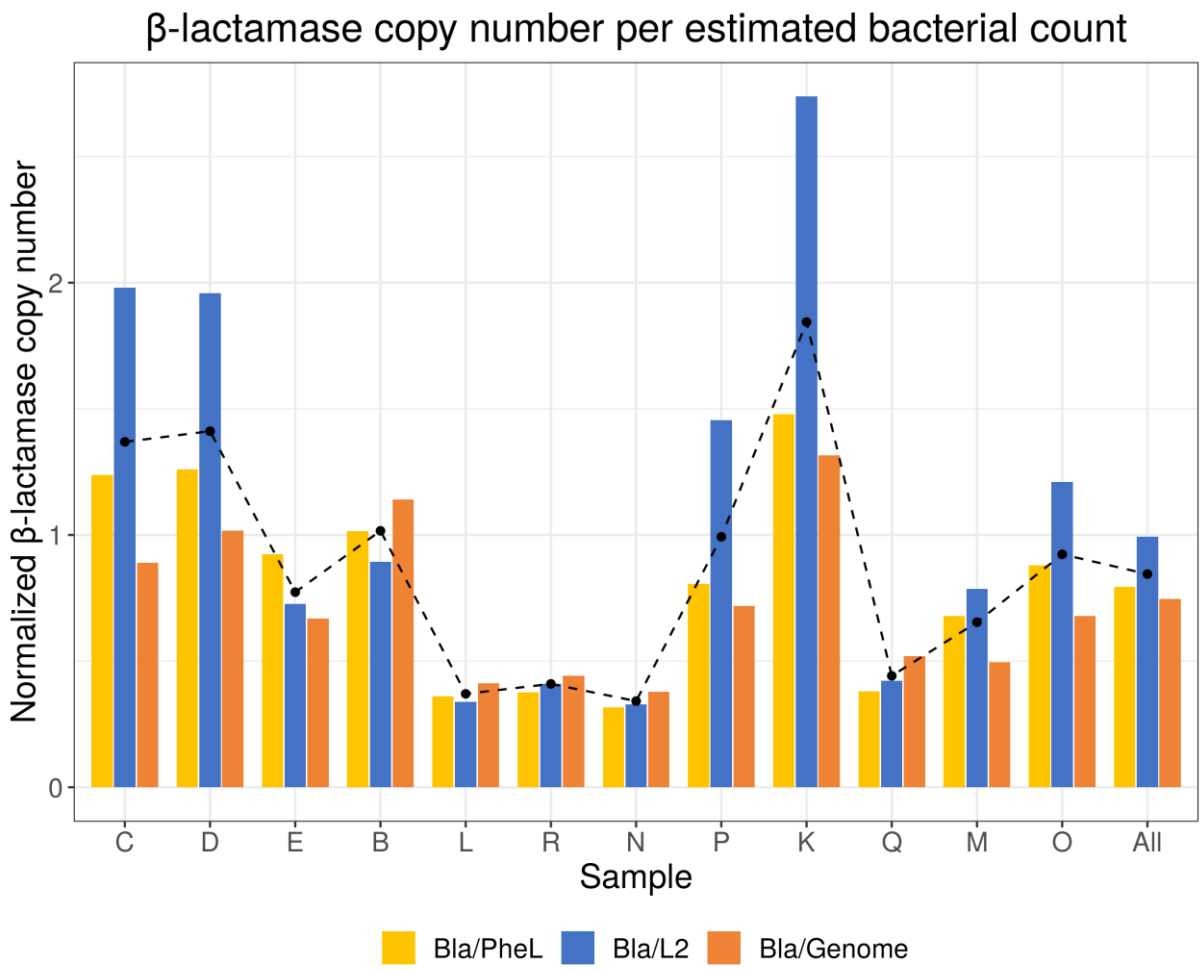


Figure 10

Legend: *β-lactamase gene (Bla) copy number estimated by comparison with the abundance of two single-copy bacterial genes: phenylalanyl-tRNA ligase (PheL) (yellow), ribosomal protein L2 (blue), and the number of equivalent bacterial genomes sequenced (orange). The average of these three determination is indicated by a black dotted line. Across all samples and all references the global average β-lactamase gene copy number per bacteria is 0.87 (95% confidence interval: [0.56 – 1.17]). See detailed statistics in Table S1.*

This copy number estimate is obtained by simply dividing the β-lactamase gene count by the number of two different universal bacterial single-copy genes, the Phenyl-alanine tRNA ligase, and the Ribosomal protein L2 (Fig. 10).

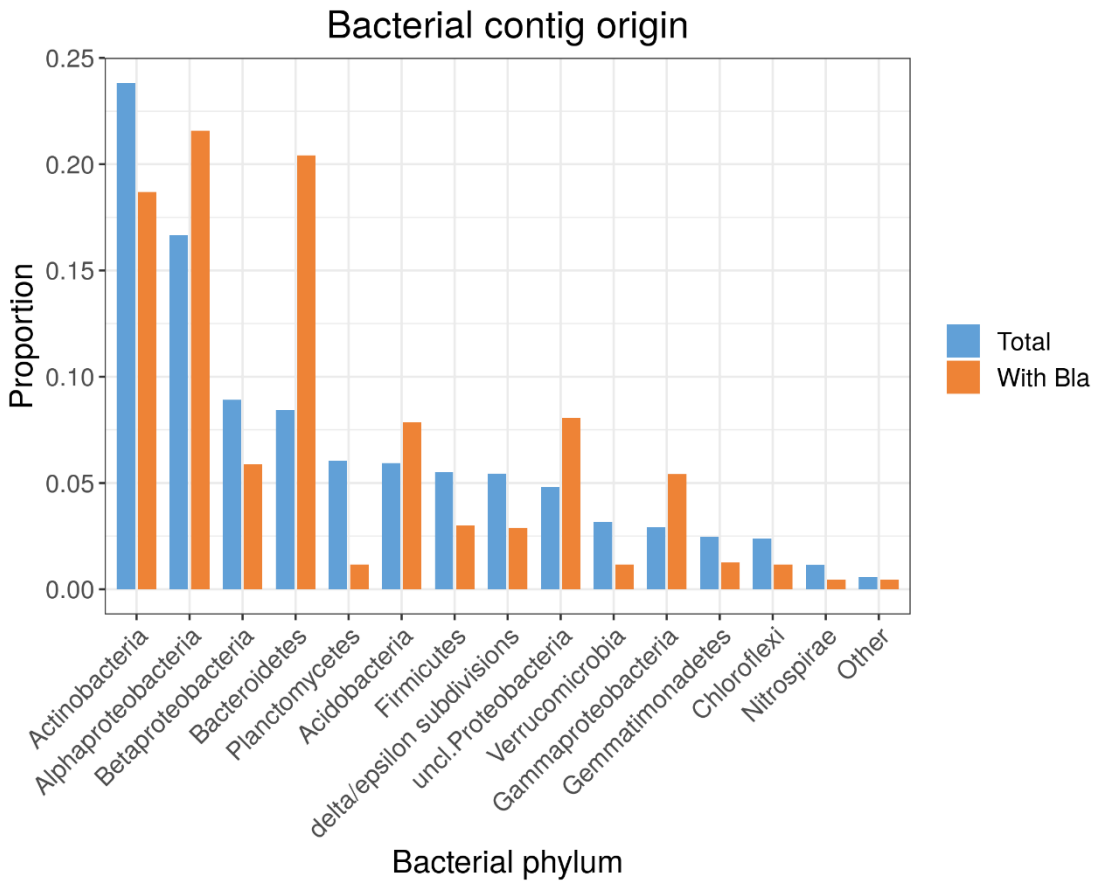
The unexpected large proportion (close to 100%) of bacteria carrying at least one β-lactamase gene indicated by the above method, was independently confirmed by calculating the ratio of the number of β-lactamases detected by the number of equivalent bacteria genomes corresponding to the total contig length for each sample. We used 3.87 Mb as the average bacterial genome sizes (68). This determination of the average β-lactamases gene copy number (0.71, 95% confidence interval: [0.51-0.]) was in good agreement with the value computed in reference to the Phenyl-alanine tRNA ligase (0.81, 95% confidence interval: [0.57-1.06], t-test p>0.45), and slightly below the one in reference to the Ribosomal protein L2 (1.11, 95% confidence interval: [0.61-1.6], t-test p>0.11).

Although a visual inspection of Fig. 10 suggests that surface cryosoils (samples C, D, E) have a higher prevalence of β-lactamase-encoding bacteria (e.g. mean: 0.87 Bla/Genome, 95% confidence interval: [0.47-1.27],) than subsurface ones (Mean = 0.66, 95% confidence interval: [0.42 -0.89]), the difference is yet not statistically significant (t-test p>0.33), due to larger fluctuations in the subsurface samples. We noticed that the lowest prevalence of β-lactamase-encoding bacteria correspond to samples from the same boreholes (L, R, N, and Q, M), suggesting that it reflects an ecological reality.

Despite the above fluctuation in the average β-lactamase copy number (0.33 (N)-1.84 (K), Fig.10), the relative proportions of the 4 enzyme classes remains fairly stable across samples. Class C [52-84%] is always the most represented, followed by class B and A (respectively 16% and 14% in average), then Class D (10%) (Fig. 9).

Finally, we explored to what extent the β-lactamase genes could be unevenly distributed among the various bacterial phylum identified in the metagenomics data. For instance, one given group of bacteria could concentrate most of β-lactamase genes eventually located onto a plasmid in very high copy number. We thus simply compared the relative abundance of each bacterial phylum within the set of all cognate contigs versus those exhibiting a β-lactamase gene. Although the two distributions (Fig. 11) are far from being identical, they remain significantly correlated (Pearson coefficient > 0.8) pointing the same phyla as predominant in both distributions (Alphaproteobacteria, Bacteroidetes, Actinobacteria). Thus, the high frequency of β-lactamase genes detected in the sampled cryosoils is not a property of a specific group of bacteria, but is shared among the very different phyla constituting these microbiomes.

Figure 11



***Legend**: Side-by-side comparison of the proportions of total contigs from a given bacterial phylum with that of β-lactamase-containing genes. These proportions are globally correlated (Pearson correlation coefficient > 0.8), indicating that β-lactamases genes do not originate from a single dominant phylum. Bacteroidetes appears the most enriched phylum in β-lactamase-encoding species.*

We set up to further validate the above findings by two additional approaches. First we extended our analysis to contigs smaller than 5k, but still large enough to statistically encompass a complete β-lactamase coding region (about a thousand bp). The much larger number of contigs in this size range (more than 2 million, representing 1539 equivalent bacterial genomes) could thus provide an independent estimate of the proportion of β-lactamase-encoding bacteria in our sample. We summarized the results in Fig. S6. They are consistent with the β-lactamase prevalence estimated from larger contigs (mean=0.65, 95% confidence interval: [0.45-0.85]) (Table S2), although a fraction of the β-lactamase genes (about 1000 bp in size) is expected to be missed when they only partially overlap with short contigs.

In a second independent control experiment, we applied our contig assembly protocol to 3 previously published metagenomic datasets generated from deep permafrost from Yedoma deposits located on the Bykovsky Peninsula (coordinates: 72.0069 N 129.1033 E, depth: 24 m, dated >50,000 y), the Gydanskiy Peninsula (coordinates: 72.3741 N 78.6116 E , depth: -4.5 m, dated > 34,000 y), and the Omolon River mouth (coordinates: 68.7166 N 158.9000 E , depth: -26 m, dated 25,000 y) (69)(Table S3). These contigs were then searched for the presence of encoded β-lactamases, with the exact same bioinformatics protocol used onto our own data. Although these datasets were approximately 15 times smaller than ours (in term of total sequenced bp) and assembled in an even lesser number of contigs (Table S3) the results (although subject to larger statistical fluctuations) were again in agreement with the level of prevalence of β-lactamases computed from of our large-scale analyses (Table 3).

Table 3: Analysis of 3 independent datasets

Dataset Accession	site	Total bp in >5kb contigs	Bla	Bact genome	Bla/genome	L2	Bla/L2	PheL	Bla/PheL	Average
SRX2163490	Bykovskiy P.	33,997,197	4	8.8	0.45	10	0.4	11	0.36	0.4
SRX2163491	Omolon River	6,278,141	1	1.6	0.62	2	0.5	2	0 .5	0.54
SRX2163492	Gydanskiy P.	61,197,557	8	15.8	0.51	6	1.3	18	0.61	0.81
	Total	101,472,895	13	26.22		18		31		0.52

Discussion

An unexpectedly variable DNA content

We found the quantities of DNA recovered from the various cryosoils to be highly variable, from more than 10 µg/g to 0.02/g, thus a factor of 500 (Table 2). Most of the low yielding samples corresponded to ancient (subsurface) permafrost layers, although without a strict correlation with age. The sample from the Stanchikovskiy Yar (B) exhibited an exceptionally high DNA contents. However, once purified, the recovered DNA exhibited similar sequencing yields (i.e. number of read/ng) and assembled into contigs of similar sizes in average (Table 2). We interpret this result as suggesting that the lower efficiency in DNA recovery from the ancient samples mainly reflects a lower global density of microorganisms rather than DNA degradation alone (at least to a point precluding its sequencing). Such a result is important because it supports the hypothesis that ancient genes carried by the ancient DNA could be recycled within contemporary microorganisms though transformation.

Bacteria dominate DNA contents

As expected, bacteria constituted most of the DNA sources (Fig. 2). The sample displayed a large variability in the composition of their bacterial populations, as already visible from their different distribution of G+C% (Fig. 1). Such variability was reflected at the taxonomical phylum level, where the globally dominant taxa (proteobacteria, actinobacteria, bacteroidetes, acidobacteria and firmicutes) were detected in highly variable relative proportions, including among samples from different depth in the same borehole (Fig. 3). This suggests the absence of a significant mixing between the distinct microbiomes of the permafrost layers. However, the taxonomical diversity of these microbiomes below the phylum level (and the corresponding metabolic diversity) precludes the interpretation of their composition in terms of past properties of the corresponding horizon. Such intra-phylum diversity is illustrated for the Proteobacteria in Fig. 4. There is a growing opinion that taxonomy alone is insufficient to inform us on the complexity of soil processes (70) given the already high diversity of metabolisms and life-styles encountered in a given phylum (such as the proteobacteria). Globally, the phyla found to dominate the microbiomes in our cryosoil and permafrost samples do not differ from those already reported as most abundant in previous analyses of temperate soils (71) or permafrost (72, 73). Those are proteobacteria (37%), actinobacteria (22%), bacteroidetes (12%), acidobacteria (6%), firmicutes (5%) and plantomycetes (5%) (Fig. 3). However, even the most common phyla do exhibit very different proportions between samples: 74% to 16.2% for proteobacteria, 52.4% to 2.1% for actinobacteria, 34.5% to 1.9% for bacteroidetes (Fig. 3). We noticed that the same phyla dominated both in the surface cryosoils and the permafrost sample except for the firmicutes that are almost uniquely found in subsurface samples (>1.1% at the surface, up to 24.3% in permafrost). However, this could be a bias of our limited number of samples. If not, it could be related to specific properties of these bacteria, such as endospore formation and anaerobiosis (e.g. Clostridia).

Archeons are less abundant than expected

The production of methane, a potent greenhouse gas, is a frequently cited global-warming threat as associated with long-term thawing permafrost (42, 74, 75 and references herein). In the future, part of this flux might originate from the reactivation of methanogenic archaeons trapped in deep permafrost layers. In this respect, the very low abundance of the archaeal microbial population in most samples came as a surprise. Out of 9 analysed permafrost samples only 4 exhibited more than 1% amount of methanogenic archaeons (Bathyarcheota or Methanomicrobia)(Fig. 5). Also interesting was the fact that 3 of these samples originated from various depth (respect. 12 m, 16m and 19m) of a the same borehole. This suggests that the presence of a sizable archaeal population is not a universal property of permafrost (even within the Yukechi site), and that its detailed composition is also variable and dependent of paleoenvironmental conditions during soil formation, as previously noticed (75). Because of their apparent scarcity, the contribution of revived ancient frozen methanogens to methane production and global warming might be less than anticipated.

A surprising scarcity of Eukaryotes

An unexpected result 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). This result does not seem consistent with 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. The lack of eukaryotic contigs can be explained in several ways, in the first place by the larger size of eukaryotic genomes. For a given amount of sequenced DNA (thus a given number of cells), a larger genome will result in a smaller coverage. However, statistically, the average size of the contigs is directly related to the coverage value. For a contig size of 5 kb, this value is ≈7, according to our assembly parameters (76). It is therefore possible that many eukaryotes do not reach such threshold. This handicap is reinforced if the proportion of eukaryotic cells is much lower than that of bacteria, as it is for archaeons. The opposite hypothesis, though less likely, could be that the eukaryotic diversity is higher, thus reducing the coverage for each of them. Fragmentation of eukaryotic coding regions by introns, and their lower density due to large intergenic regions may also strongly decrease the probability of their identification by similarity searches in protein databases. Contigs without taxonomic assignment could therefore represent unaccounted eukaryotic contigs, but their proportion remains negligible (Fig. 2, “no taxonomy” contigs). Finally, it is also possible that the DNA of eukaryotic origin is rapidly recycled into microbial DNA during the seasonal decomposition process that takes place in the superficial active layer before being perennially frozen in the deeper permafrost.

A low proportion of viral contigs

Contigs identified as viral origin are by far the least numerous in our analysis (Fig. 2), in particular bacteriophages (Caudovirales in Fig. 8). This is an unexpected result given that

bacteria form the vast majority of microorganisms present in our samples and are generally the hosts of a wide variety of bacteriophages. One possibility is that phage populations are in fact very large, but too diverse, thus preventing the assembly of sufficiently long contigs (> 5 kb) to be formed and counted. In addition, with their very few common genes and high sequence variability, bacteriophages are notoriously difficult to identify from partial metagenomic sequences (77). They may therefore be included within contigs without taxonomic classification. Their frequent exchange of genes with different bacterial hosts can also cause them to be classified as chimeric contigs (Fig. 2). However, even considering these adjustments, their proportion would never exceed more than a few %. Finally, it is also possible that the genomes of bacteriophages, like that of other viruses, are under-represented because of the use of an inadequate DNA extraction protocol.
Regarding eukaryotic viruses, our results are nevertheless less surprising, since they show a predominance for families of viruses infecting protozoa whose presence in ancient permafrost is known (33, 34). In addition, some members of these viral families have been reactivated (23, 24). We noticed that the phylum Amoebozoa, known to be that of the hosts of families of large DNA viruses (Mimivirus, Pithovirus, Fig. 8) is among the 3 predominant eukaryotic clades (samples R and N, in particular).
Finally, on an optimistic note, we detected a very low proportion of viruses possibly pathogenic for animals (such as Asfarviridae and Iridoviridae), a result which puts into perspective the potential risk of viral infections triggered by the awakening of "zombie" viruses released by the thawing of permafrost. This danger could still remain occasionally near the corpses of humans (21, 22) or animals (19, 20, 78, 79) well preserved in permafrost. Nevertheless, we also have to remember that none of the viruses we managed to isolate from ancient permafrost and surface cryosoil were present in large enough abundances to achieve sizable contigs from metagenomics data (24, 80).

Thawing permafrost as an abundant source of antibiotic resistance genes

The presence of antibiotic-resistant bacteria carrying various types of β-lactamases (A, B, C and D, according to Ambler's classification (67)) has been previously reported in pristine Arctic and Antarctic surface cryosoils and ancient permafrost (61, 63, 81-84). 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, in deep permafrost).
In this work, we evaluated this proportion using two independent approaches. In the first, we compared the frequency of β-lactamase genes identified in our metagenomics data with those of two reference genes known to be present in a single copy in bacterial genomes. The second approach consisted in comparing the number of identified β-lactamase genes to the theoretical number of bacterial genomes equivalent to the total length of contig sequences. In both cases, our results indicated that more than half of the bacteria in our samples encode a β-lactamase in average (0.87, 95% confidence interval: [0.56 – 1.17]).

Similar high proportion of β-lactamase-encoding bacteria were computed from our own dataset extended to smaller contigs (Fig. S6, Table S2), and more surprisingly from metagenomics datasets generated from deep permafrost collected from remote locations (69) and different horizons. This suggests that the high prevalence β-lactamase-encoding bacteria might be a common property of cryosols, relatively independent from their detailed composition (although this would need to be confirmed by more metagenomic studies explicitly designed for this purpose).

We looked to confront our results to the existing literature. Unfortunately, we found very few studies allowing our findings to be compared with previous experimental studies. Indeed, almost all investigations of various environmental "resistomes" were performed on populations of culturable bacteria preselected for their antibiotic resistance. A few others are limited to a pre-defined bacterial species (e.g. *Escherichia coli*) (85). We could found one study analysing the proportion of β-lactam-resistant bacteria within the global bacterial population of different agricultural soils, untreated or amended with manure (86). This proportion was found to vary from 0.67% to 7.4% in untreated vs. manure-amended soils. Most of these resistant bacteria presumably encoded a β-lactamase, although to a much lesser proportion than found in our cryosoil samples.

Another shotgun metagenomics analysis of the distribution of antibiotic resistance genes (ARG) 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 would approximately correspond (assuming 3850 as an average gene content per bacteria) to 17% of soil bacterial encoding at least one ARG (83). However, 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.

To the best our knowledge, the high proportion of β-lactamase-encoding bacteria revealed by our metagenomics analysis of 12 diverse cryosoil samples (and 3 previously published datasets) (69), is significantly greater than what has been previously reported for other pristine environments, undisturbed by anthropogenic activities. As the function of β-lactamase in native soil microbial community is not known, our results should stimulate the building of new hypotheses. Certain soil microorganisms produce antibiotic β-lactams, including members of the Actinobacteria (e.g. Streptomyces species) (87). We noticed that Actinobacteria are among the most abundant in the cryosoil microbiomes (Fig. 3). Among eukaryotes, Fungi (also abundant in our samples, Fig. 7) have the metabolic resources to produce many complex lactam-containing compounds. They are the original source of two foundational β-lactam antibiotics: penicillin and cephalosporin (88). The higher than usual abundance β-lactamase-carrying bacteria in the cryosoils might thus reflect a particular intense war between microbes occurring during the specific perennial decomposition and freezing/thawing processes characteristic of Polar Regions. However, antibiotic resistance genes may play roles outside of the simplistic "war" paradigm. At subminimal inhibitory concentrations, antibiotics can modulate bacterial gene expression thus acting as signalling

molecules rather than weapons. In such context, β-lactamases might disrupt such signalling and interfere with quorum sensing (89).

Our study indicated the presence of the four classes of β-lactamases, 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 (90). Class B β-lactamases are metallo-enzymes 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 (90). Class C β-lactamases derives from the ampC gene carried on the genome of many *Enterobacteriaceae.* Variants of these enzyme are known to reduce sensitivity to carbapenems (90). Finally, class D β-lactamases, also known as oxacillinases (OXAs) evolved from degrading an extended spectrum of cephalosporins to hydrolyze carbapenems (90). However, from their sequences alone (except for near 100% identical residues) it is impossible to predict the substrate range and the potential clinical risk of the diverse β-lactamases gene identified in our data set.

Prompted by our previous discovery of “living” viruses in Siberian permafrost (as well as several others yet to be published), we initiated this large-scale metagenomic investigation at various sites to refine our assessment of the public health risk posed by the release of unknown ancient viruses in the context of global warming and the accelerated thawing of permafrost. We found that viruses constituted the smallest part of the various cryosoil microbiomes (Fig. 2), consisting mostly of protozoan-infecting viruses and bacteriophages. Except in the immediate proximity of frozen animal or human remains, the risk of viral infections is thus probably negligible for us.

However, this study revealed a new potential danger, consisting in the unexpected large abundance of β-lactamases genes carried in the genomes (or plasmids) of a large phyletic diversity of cryosoil bacteria. The DNA from these bacteria, dead, alive, or in a frozen dormancy, thus constitutes an immense reservoir of historical antibiotic resistance genes the transfer of which to contemporary human bacterial pathogens 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, and in particular rapid Arctic warming, thawing permafrost will behave as a resistance gene-flavoured ice cream to be enjoyed without moderation by all bacteria passing by.

Funding

This work was supported by France Génomique Grant ANR-10-INSB-01-01 to J.-M. C and CNRS PRC research grant (PRC1484-2018) to C.A. E C-F was supported by a PhD grant (DGA/DS/MRIS) #2017 60 0004. GG and JS were funded by ERC PETA-CARB (#338335) and the HGF Impulse and Networking Fund (ERC-0013).

Acknowledgements

We are deeply indebted to our volunteer collaborator, Alexander Morawitz, for collecting the Kamchatka soil samples. We thank the PACA Bioinfo platform for computing support. 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.

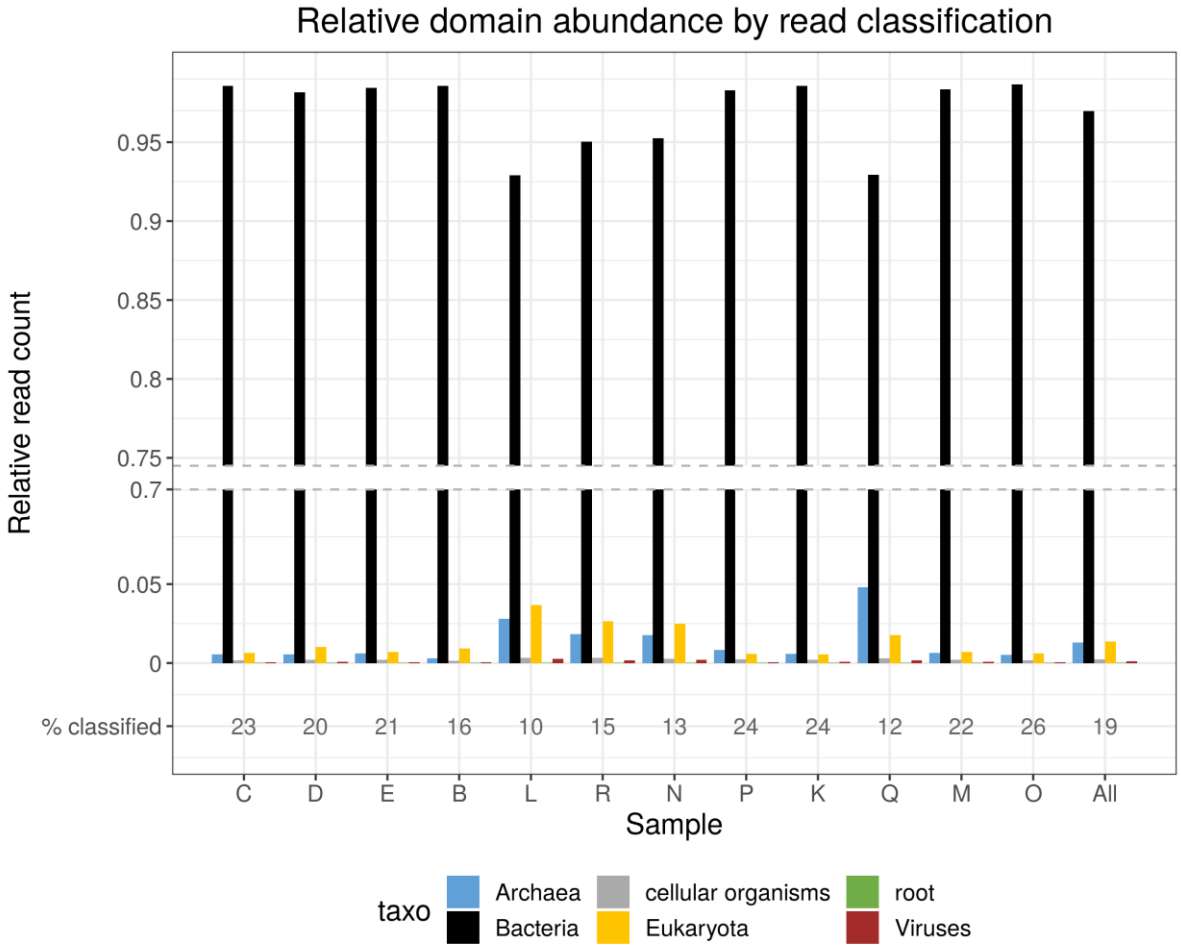
Contributions

JMC and CA designed the study, SR and SS performed the bioinformatics analyses, EC processed the sample and performed DNA extraction, KL supervised DNA sequencing, GG, JS, and ANF collected the deep core samples from the Yukechi site. All authors contributed to the manuscript writing.

Supplementary Figure S1

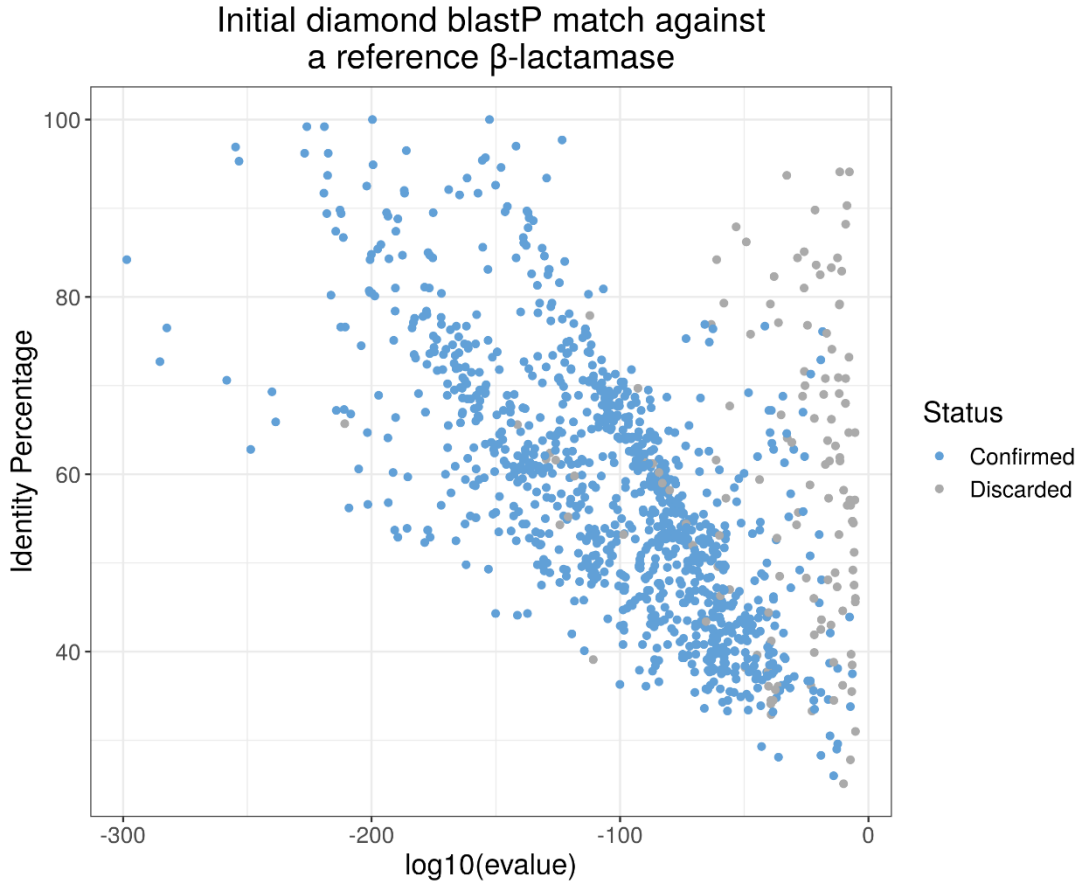
Legend. *Overview of sampling locations (left map) with details on the sampling sites from Yukechi (upper right) and Kamchatka (lower right)*

Supplementary Figure S2



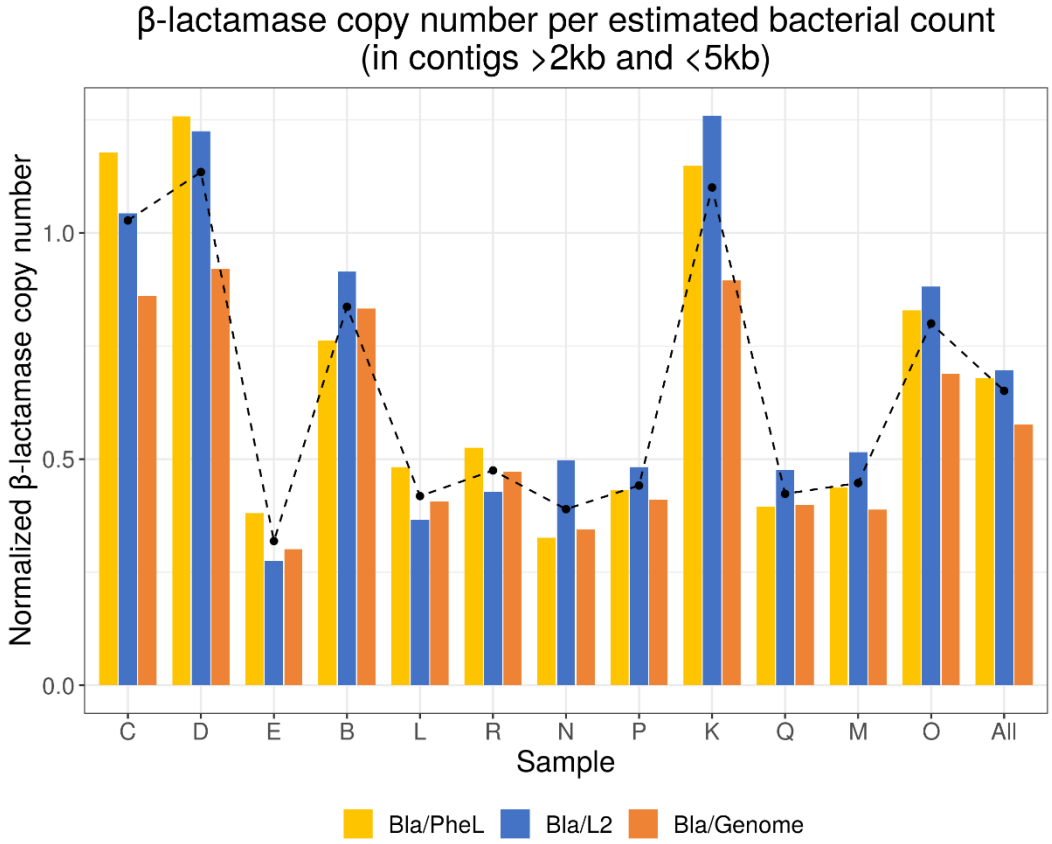
***Legend**: Relative distributions of organism types identified in the various samples, using a read-level algorithm (Kraken). Notice the small percentage (% classified) of used reads.*

Supplementary Figure S3



Legend: Similarity (% identical residues) and matching P-value of predicted ORFs with a best match with a β-lactamase in the Refseq database using Diamond. ORFs were then discarded or confirmed based on size and the presence of a *bona fide* β-lactamase domain (See Material and methods). The bulk of predicted sample β-lactamases are in the unambiguous [30%-90%] identity range.

Supplementary Figure S4



***Legend**: β-lactamase gene (Bla) copy number estimated by comparison with the abundance of two single-copy bacterial genes: phenylalanyl-tRNA ligase (PheL) (Blue), ribosomal protein L2 (orange), and the number of equivalent bacterial genomes sequenced (grey). The average of these 3 determination is indicated by a black dotted line. Across all samples, the global average β-lactamase gene copy number per bacteria is 0.65 (95% confidence interval: [0.45 – 0.84]). A very similar picture is obtained when applying a less stringent criteria not requesting each sequence to encompass the relevant complete β-lactamase domain (Fig. 10).*

Supplementary Table S1. Analysis of contigs larger than 5 kb

Sample	Bla	L2	PheL	Bac. genome	A	B	C	D	Bla/L2	Bla/PheL	Bla/Bact	Average	N contig >5kb	Total length (bp)
C	99	50	80	111,6	14	16	59	10	1,98	1,24	0,89	1,37	42,130	431,858,365
D	233	119	185	230,4	45	45	124	19	1,96	1,26	1,01	1,41	78,167	891,601,631
E	24	33	26	36,5	0	2	18	4	0,73	0,92	0,66	0,77	16,641	141,332,690
B	68	76	67	59,8	6	13	40	9	0,89	1,01	1,14	1,02	15,293	231,289371
L	44	121	114	105,1	2	4	37	1	0,36	0,39	0,42	0,39	34,431	406,791,013
R	41	100	109	108,5	4	4	29	4	0,41	0,38	0,38	0,39	35,616	420,006,247
N	26	76	79	80,6	5	3	14	4	0,34	0,33	0,32	0,33	28,197	312,114,276
P	112	77	139	157,5	5	21	81	5	1,45	0,81	0,71	0,99	64,903	609,516,216
K	167	61	113	128,4	17	21	118	11	2,74	1,48	1,30	1,84	49,767	497,022,492
Q	79	182	202	153,6	1	8	60	10	0,43	0,39	0,51	0,45	38,175	594,545,046
M	70	89	103	144,9	6	12	48	4	0,79	0,68	0,48	0,65	58,035	560,831,523
O	109	90	124	161,2	10	16	68	15	1,21	0,88	0,68	0,92	63,958	623,96,6161
All	1072	1074	1341	1478	115	165	696	96	1.11	0.81	0.71	0.88	525,313	5,720,875,031

Supplementary Table S2. Analysis of 2 kb < contigs <5 kb

Sample	Bla	L2	PheL	Bact. Genome	Bla/L2	Bla/PheL	Bla/Bact	Average	N contig 2kb>L>5kb	Total bp
C	128	122	108	148,2	1,05	1,19	0,86	1,03	200,802	573,410,069
D	191	155	151	207,3	1,23	1,26	0,92	1,14	275,382	802,389,205
E	26	94	68	87,4	0,28	0,38	0,30	0,32	120,522	338,460,150
B	23	25	30	27,5	0,92	0,77	0,84	0,84	35,731	106,410,805
L	32	87	66	82,6	0,37	0,48	0,39	0,41	108,816	319,699,512
R	37	86	70	91,4	0,43	0,53	0,40	0,45	121,633	353,910,437
N	21	42	64	74,1	0,50	0,33	0,28	0,37	98,513	286,741,299
P	80	165	184	195,6	0,48	0,43	0,41	0,44	258,914	756,843,977
K	157	124	136	176,6	1,27	1,15	0,89	1,10	237,762	683,392,792
Q	33	69	83	85,4	0,48	0,40	0,39	0,42	112,563	330,646,859
M	69	133	157	181,4	0,52	0,44	0,38	0,45	241,064	701,920,384
O	125	141	150	181,4	0,89	0,83	0,69	0,80	239,974	701,853,105
All	922	1243	1267	1539	0.70	0.68	0.56	0.65	2,051,676	5,955,678,594

Supplementary Table S3. Processing of the datasets from Vishnivetskaya et al. (69)

Accession	Source	Total bp after cleaning	N contig > 5kb	Average size[1]	Median size[2]	Median coverage[2]
SRX2163490	Bykovsky Peninsula	18,845,162,025	4382	7758	1379	9.0
SRX2163491	Gydan Peninsula	20,905,522,718	836	7509	1318	44.
SRX2163492	Omolon River	26,577,180,575	9136	6698	1395	6.95

1: computed on the contig > 5 kb ; 2: computed on all contigs>1 kb

References

1 Morens DM, Daszak P, Markel H, Taubenberger JK. Pandemic COVID-19 Joins History's Pandemic Legion. *mBio* 2020; **11**: e00812-20.

2 Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**: 270-273.

3 Marini G, Calzolari M, Angelini P, et al. A quantitative comparison of West Nile virus incidence from 2013 to 2018 in Emilia-Romagna, Italy. *PLoS Negl Trop Dis* 2020; **14**: e0007953.

4 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.

5 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.

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

7 Keita MB, Hamad I, Bittar F. Looking in apes as a source of human pathogens. *Microb Pathog* 2014; **77**: 149-154.

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

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

10 Plowright RK, Parrish CR, McCallum H, et al. Pathways to zoonotic spillover. *Nat Rev Microbio*l 2017; **15**: 502-510.

11 Revich BA, Podolnaya MA Thawing of permafrost may disturb historic cattle burial grounds in East Siberia. *Global Health Action* 2011; **4**: 1.

12 Revich B, Tokarevich N, Parkinson AJ. Climate change and zoonotic infections in the Russian Arctic. *Int J Circumpolar Health* 2012; **71**: 1.

13 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**: 1.

14 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**: 1.

15 Cohen J, Screen J, Furtado J, et al. Recent Arctic amplification and extreme mid-latitude weather. *Nature Geosci* 2014 **7**: 627–637.

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

17 Turetsky MR, Abbott BW, Jones MC, et al. Permafrost collapse is accelerating carbon release. *Nature* 2019; **569**: 32-34.

18 Fuchs M, Ingmar N, Strauss J, et al. Rapid Fluvio-ThermalErosion of a Yedoma Permafrost Cliff in the Lena River Delta. *Front Earth Sci* 2020; **8**: 336.

19 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.
20 Hueffer K, Drown D, Romanovsky V, Hennessy T. Factors Contributing to Anthrax Outbreaks in the Circumpolar North. *Eco health* 2020; **17**: 174-180.
21 Biagini P, Thèves C, Balaresque P, et al. Variola virus in a 300-year-old Siberian mummy. *N Engl J Med* 2012; **367**: 2057-2059.
22 Crubézy E, Amory S, Keyser C, et al. Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC Evol Biol* 2010; **10**: 25.
23 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 U S A* 2014; **111**: 4274-4279.
24 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 U S A* 2015; **112**: E5327-E5335.
25 Hinsa-Leasure SM, Bhavaraju L, Rodrigues JL, et al. Characterization of a bacterial community from a Northeast Siberian seacoast permafrost sample. *FEMS Microbiol Ecol* 2010; **74**: 103-113.
26 Graham DE, Wallenstein MD, Vishnivetskaya TA, et al. Microbes in thawing permafrost: the unknown variable in the climate change equation. *ISME J* 2012; **6**: 709-712.
27 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**: 400-414.
28 Gilichinsky D, Microbial life in Permafrost: a historical review. *Permafrost Periglacial Process* 1995; **6**: 243-250.
29 Johnson SS, Hebsgaard MB, Christensen TR, et al. Ancient bacteria show evidence of DNA repair. *Proc Natl Acad Sci U S A* 2007; **104**: 14401-14405.
30 Willerslev E, Hansen AJ, Poinar HN. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 2004; **19**: 141-147.
31 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 U S A* 2012; **109**: 4008-4013.
32 Kochkina G, Ivanushkina N, Ozerskaya S, et al. Ancient fungi in Antarctic permafrost environments. *FEMS Microbiol Ecol* 2012; **82**: 501-509.
33 Shmakova L, Bondarenko N, Smirnov A. Viable Species of Flamella (Amoebozoa: Variosea) Isolated from Ancient Arctic Permafrost Sediments. *Protist* 2016; **167**: 13-30.
34 Malavin S, Shmakova L. Isolates from ancient permafrost help to elucidate species boundaries in Aca*nthamoeba castellanii* complex (Amoebozoa: Discosea). *Eur J Protistol* 2020; **73**: 125671.
35 Schirrmeister L, Froese DG, Tumskoy V, et al. Yedoma: Late Pleistocene ice-rich syngenetic permafrost of Beringia. S.A. Elias (Ed.), Encyclopedia of Quaternary Sciences, Elsevier, Amsterdam (2013), pp. 542-552.

36 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.
37 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.
38 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**: 3469–3488,
39 Turetsky MR, Abbott BW, Jones MC, et al. Carbon release through abrupt permafrost thaw. *Nat Geosci* 2020; **13**: 138–143.
40 Fedorov A, Konstantinov P. Observations of surface dynamics with thermokarst initiation, Yukechi site, Central Yakutia. *Proceedings of the 8th International Conference on Permafrost*, Zurich, Switzerland, 2003; pp. 239–243.
41 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:** 3797–3814.
42 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 Biology* 2021; doi: 10.1111/gcb.15566
43 Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
44 https://jgi.doe.gov/data-and-tools/bbtools/
45 Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; **31**: 1674-1676.
46 Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; **9**: 357-359.
47 Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 2010; **38**: e132.
48 Li W, et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res* 2021; **49**: D1020-D1028.
49 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 2015; **12**: 59-60.
50 Wood E, Lu Jr, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019; **20**: 1-13.
51 Jones P, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236-1240.
52 Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; **22**: 1658-1659.
53 Lu S, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020; **48**: D265-D268.

54 Pellow D, Mizrahi I, Shamir R. PlasClass improves plasmid sequence classification. *PLoS Comput Biol* 2020; **16**: e1007781.
55 Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 2018; **46**: e35.
56 Evans PN, Parks DH, Chadwick GL, et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 2015; **350**: 434-438.
57 Hatzenpichler R. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol* 2012; **78**: 7501-7510.
58 Behera P, Mohapatra M, Kim JY, Rastogi G. Benthic archaeal community structure and carbon metabolic profiling of heterotrophic microbial communities in brackish sediments. *Sci Total Environ* 2020; **706**: 135709.
59 Davis JL, et al. Ground penetrating radar surveys to locate 1918 Spanish flu victims in permafrost. *J Forensic Sci* 2000; **45**: 68-76.
60 Göller PC, Haro-Moreno JM, Rodriguez-Valera F, Loessner MJ, Gómez-Sanz E. Uncovering a hidden diversity: optimized protocols for the extraction of dsDNA bacteriophages from soil. *Microbiome* 2020; **8**: 17.
61 D'Costa VM, et al. Antibiotic resistance is ancient. *Nature* 2011; **477**: 457-461.
62 Finley RL, et al. The scourge of antibiotic resistance: the important role of the environment. *Clin Infect Dis* 2013; **57**: 704-710.
63 Perron GG, et al. Functional characterization of bacteria isolated from ancient arctic soil exposes diverse resistance mechanisms to modern antibiotics. *PLoS One* 2015; **10**: e0069533.
64 Wright GD. Environmental and clinical antibiotic resistomes, same only different. *Curr Opin Microbiol* 2019; **5**1: 57-63.
65 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**: 568-585.
66 Cantón R, et al. Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. *Clin Microbiol Infect* 2012; **18**: 413-431.
67 Ambler RP. The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 1980; **289**: 321-331.
68 diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiol Mol Biol Rev* 2017; **81**: e00019-17.
69 Vishnivetskaya T, Spirina E, Shmakova L, Tutukina M, Li Z, Wu X, Rivkina E. Metagenomes from Late Pleistocene Ice Complex Sediments of the Siberian Arctic. *Microbiol Resour Announc* 2019; **8**: e01010-19.
70 Baldrian P. The known and the unknown in soil microbial ecology. *FEMS Microbiol Ecol* 2019; **95**. doi: 10.1093/femsec/fiz005.
71 Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* 2006; **72**: 1719-1728.
72 Tripathi BM, 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

73 Xue Y, Jonassen I, Øvreås L, Taş N. Bacterial and Archaeal Metagenome-Assembled Genome Sequences from Svalbard Permafrost. *Microbiol Resour Announc* 2019; **8**: e00516-19.
74 Masyagina OV, Menyailo OV. The impact of permafrost on carbon dioxide and methane fluxes in Siberia: A meta-analysis. *Environ Res* 2020; **182**:109096.
75 Holm S, et al. Methanogenic response to long-term permafrost thaw is determined by paleoenvironment. *FEMS Microbiol Ecol* 2020; **96**: fiaa021. doi: 10.1093/femsec/fiaa021.
76 Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 1988; **2**: 231-239.
77 Al-Shayeb B, et al. Clades of huge phages from across Earth's ecosystems. *Nature* 2020; **578:** 425–431.
78 Fisher DC et al. Anatomy, death, and preservation of a woolly mammoth (Mammuthus primigenius) calf, Yamal Peninsula, northwest Siberia. *Quat. Int.* 2012; **255**: 94–105.
79 Papageorgopoulou, C.; Link, K.; Rühli, F.J. Histology of a woolly mammoth (Mammuthus primigenius) preserved in permafrost, Yamal Peninsula, Northwest Siberia. *Anat. Rec.* 2015; **29**: 1059–1071.
80 Christo-Foroux E, et al. Characterization of *Mollivirus kamchatka*, the First Modern Representative of the Proposed Molliviridae Family of Giant Viruses*. J Virol* 2020; **94**: e01997-19.
81 Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* 2009; **3**: 243-251.
82 Petrova M, Kurakov A, Shcherbatova N, Mindlin S. Genetic structure and biological properties of the first ancient multiresistance plasmid pKLH80 isolated from a permafrost bacterium. *Microbiology* 2014; **160**: 2253-2263.
83 Van Goethem MW, et al. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* 2018; 6:40.
84 Kashuba E, et al. Ancient permafrost staphylococci carry antibiotic resistance genes. *Microb Ecol Health Dis* 2017; **28**: 1345574.
85 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**: 1181-1197.
86 Udikovic-Kolic N, Wichmann F, Broderick NA, Handelsman J. Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proc Natl Acad Sci U S A* 2014; **111**: 15202-15207.
87 Ogawara H. Self-resistance in Streptomyces, with Special Reference to β-Lactam Antibiotics. *Molecules* 2016; **21**: 605.
88 Brakhage AA, Thön M, Spröte P, Scharf DH, Al-Abdallah Q, Wolke SM, Hortschansky P. Aspects on evolution of fungal beta-lactam biosynthesis gene clusters and recruitment of trans-acting factors. *Phytochemistry* 2009; **70**: 1801-1811.

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

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