

1 **Metagenomic analyses of the late Pleistocene permafrost – additional tools for**
2 **reconstruction of environmental conditions**

3

4 Elizaveta Rivkina¹, Lada Petrovskaya², Tatiana Vishnivetskaya³, Kirill Krivushin¹, Lyubov
5 Shmakova¹, Maria Tutukina^{4,7}, Arthur Meyers³, Fyodor Kondrashov^{4,5,6}

6

7 ¹Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of
8 Sciences, Pushchino, Russia

9 ²Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of
10 Sciences, Moscow, Russia

11 ³University of Tennessee, Center for Environmental Biotechnology, Knoxville, USA

12 ⁴ Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG),
13 Barcelona, Spain

14 ⁵ Universitat Pompeu Fabra (UPF), Barcelona, Spain

15 ⁶Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

16 ⁷Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, Russia

17

18 *Corresponding author E-mail: elizaveta.rivkina@gmail.com*

19 Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Moscow
20 region, Russia, 142292

21

22 **Key words:** permafrost, Kolyma Lowland, methane, metagenome, methanogens,
23 methanotrophs, late Pleistocene Ice Complex, lake sediments

24 **Abstract**

25 A comparative analysis of the metagenomes from two 30,000 year-old permafrost samples,
26 one of lake-alluvial origin and the other from late Pleistocene Ice Complex sediments,
27 revealed significant differences within microbial communities. The late Pleistocene Ice
28 Complex sediments (which have been characterized by the absence of methane with lower
29 values of redox-potential and Fe^{2+} content) showed a low abundance of methanogenic
30 archaea and enzymes from both the carbon and nitrogen cycles, but a higher abundance of
31 enzymes associated with sulfur cycle. The metagenomic and geochemical analyses
32 described in the paper provide evidence that the formation of the sampled late Pleistocene
33 Ice Complex sediments likely took place under much more aerobic conditions than lake-
34 alluvial sediments.

35

1. Introduction

Permafrost, including constantly frozen sediments of the Arctic, is a unique subsurface complex environment where microorganisms retain viability over a long period of time, from thousands to millions of years (Gilichinsky and Rivkina 2011). The impact of climate change on permafrost stability has recently been discussed widely by the scientific community (Zimov et al 2006, Walter et al 2007, Anthony et al 2014). The permafrost deposits of the North-East Siberia, which did not thaw during the Holocene climatic optimum, have attracted particular interest, especially the late Pleistocene Ice Complex deposits (Yedoma Suite) that are widespread on the East Siberian coastal plains (Lena-Anabar, Yana-Indigirka, and Kolyma lowlands) (Schirrmeister et al 2011). It was found earlier that the epigenetically (consecutively) frozen sediments of both lake and marine origin (independently of age) contain biogenic methane in concentration up to 1.3 mmol kg^{-1} ($\sim 30 \text{ ml kg}^{-1}$), whereas methane was either absent or present at trace concentrations ($< 0.0002 \text{ mmol kg}^{-1}$) in samples from the sincryogenic (simultaneously frozen) late Pleistocene Ice Complex located in the Yana-Indigirka, and Kolyma lowlands (Rivkina et al 2007, Rivkina and Kraev 2008). On the contrary, Yedoma deposits found on the Tyungylyn Terrace of the Lena River (Central Yakutia) in a zone of boreal forest contained methane in concentrations up to 6,000 ppmv in pore space of permafrost (Brouchkov and Fukuda 2002). Studies conducted on Kurungnakh Island, which is situated in the southern part of the Lena River Delta showed presence of biogenic methane in Holocene sediments (up to 3.5 mmol kg^{-1}), while methane concentrations in majority of samples from late Pleistocene Ice Complex was at a detection level and only some samples from Ice Complex sediments approximately 28 K years old showed presence of methane in concentration of 1.0 mmol kg^{-1} and methane concentration increased up to 2.0 mmol kg^{-1} in sediments either younger than 16 K or older than 40 K years old (Bischoff et al 2013). The presence of biogenic

61 methane up to 0.4 mmol kg^{-1} in the Ice Complex deposits on the Gydan Peninsula located on
62 the Siberian coast of the Kara Sea between the estuaries of the Ob and Yenisei Rivers was
63 also shown during our recent research (unpublished data). These data indicate that different
64 types of Yedoma deposits are present in different parts of the Siberian Arctic. Schirrmeister
65 and co-authors pointed out “the typical Ice Complex formation consisted of several
66 concurrent cryogenic processes, including sediment accumulation and freezing, ice
67 segregation, syngenetic ice-wedge growth, sediment reworking, peat aggradation, cryosol
68 formation, and cryoturbation” (Schirrmeister et al 2013). Different ways of Yedoma
69 formation are presumably reflected in the possibility of the methane formation and its
70 preservation in Yedoma sediments. Studies reviewed above showed absence of methane in
71 the late Pleistocene Ice Complex on the territory of the Yana-Indigirka, and Kolyma
72 lowlands.

73 Anaerobic microcosm incubation of thawed permafrost samples in a carbon dioxide- and
74 hydrogen- enriched atmosphere showed methanogenic activity in epicryogenic sediments
75 only, while this process was not observed in samples from the sincryogenic late Pleistocene
76 Ice Complex (Rivkina and Kraev 2008). Similar results were obtained during experiments
77 with radioactively labeled substrates (Rivkina et al 2002, Rivkina et al 2004, Rivkina et al
78 2007), showing an absence of methanogenic activity in the sincryogenic late Pleistocene Ice
79 Complex samples with this process evidently taking place in epicryogenic sediments of lake
80 or lake-alluvial origin. Moreover, only from the latter sediments ($1.3 \text{ mmol CH}_4 \text{ kg}^{-1}$)
81 methanogenic archaea were isolated in to pure culture (Rivkina et al 2007, Krivushin et al
82 2010, Shcherbakova et al 2011). Viable methanogens were also isolated from the
83 permafrost-affected soils located in the Lena River Delta region (Morozova et al 2007,
84 Wagner et al 2013, Liebner et al 2007).

85 The specific distribution of methane and methanogenic activity in permafrost raises number
86 of questions bound to composition of microbial communities and particularly to abundance
87 and composition of methane-producing microorganisms. Certainly, the application of new
88 methodologies such as metagenomic analyses is required to estimate diversity and
89 complexity of permafrost microbial communities, and get insight into their metabolic
90 capabilities (Graham et al 2012, Jansson and Tas 2014). Until recently, the determination of
91 microbial diversity in low biomass environments, including permafrost, was problematic.
92 The microbial cell abundances in the ancient permafrost is 10–100 times lower than that in
93 the active layer samples, thereby resulting in low yields of the total community genomic
94 DNA (gDNA) (Yergeau et al 2010). However, using appropriate DNA extraction kits
95 (Vishnivetskaya et al 2014) and the whole-community genome amplification technique
96 (Yergeau et al 2010), a sufficient amount of gDNA can be obtained for next-generation
97 sequencing technologies, producing sequences on an unprecedented scale. Indeed, the first
98 metagenomic analyses of permafrost samples became available recently (Mackelprang et al
99 2011, Jansson and Tas, 2014). Specifically, the analyses of the metagenomes from active
100 layer soil and two-meter deep permafrost samples collected in the Canadian High Arctic and
101 Alaska identified signature genes responsible for hydrogenotrophic and acetoclastic
102 methanogenesis, methylotrophic methane oxidation, nitrification, and carbohydrate
103 degradation (Yergeau et al 2010, Mackelprang et al 2011).

104 Here we report results of the comparative metagenomic analyses of the two ancient
105 permafrost samples similar in age (*ca.* 30,000 years old), however of different origins
106 (epigenetically frozen lake sediments versus syncryogenic sediments from the late
107 Pleistocene Ice Complex). The aim of the study was to analyze the abundance and
108 composition of microbial communities from the late Pleistocene permafrost of different
109 origin and elucidate why the sediments of the late Pleistocene Ice Complex in the Kolyma-

Indigirka lowland do not contain methane or methanogenic activity. The second objective of the study was to compare putative metabolic capabilities of the communities and use these findings for explanation of a possible scenario of the development of low methane Yedoma sediments.

2. Materials and Methods

Sample collection and description

Samples were collected within the Kolyma-Indigirka Lowland in northeast Siberia (152-162°E, 68-72°N) during the summer field season of 2007 (Fig. 1A). Permafrost sediments were sampled using drilling equipment that operates without fluids and prevents down-hole contamination. The sampling technique was tested and described previously (Shi et al 1997). Briefly, the surfaces of the 20- to 30-cm-long cores were cleaned immediately by shaving melted layers out with an ethyl alcohol-sterilized knife and then the frozen internal part of the core was split into 5-cm-long segments; these were placed into sterile aluminum containers and kept frozen during storage in field and transportation to the Institute of Physicochemical and Biological Problems in Soil Science, Pushchino. The physicochemical characteristics of the samples are presented in Table 1. Gas samples were collected by degassing 50 g of frozen cores in a 150-mL syringe under nitrogen atmosphere as described previously (Rivkina et al 2007, Rivkina and Kraev 2008). CH₄ concentration was measured by headspace-equilibration using a gas chromatograph KhPM-4 (Moscow Company “Chromatograph”, Russia) with flame ionization detector and hydrogen as a carrier gas. The $\delta^{13}\text{C}$ of methane was analyzed on GC Combustion III Thermo Finnigan interface and Deltaplus XL mass spectrometer (Thermo Electron Corporation, Germany).

Sample IC4 corresponded to the permafrost sediment of lake origin from the floodplain of the Ambolikha River, borehole DH-4/07, depth of 22.5 m (Fig. 1B). Total carbon concentration was ~1.1% (w/w). Methane content of this sample was 1.2 mmol kg⁻¹, $\delta^{13}\text{C} = -85\text{‰}$ indicative of biogenic origin. The radiocarbon age of this sample was 30,696 \pm 394 years (J-5829) (Kraev et al 2013).

Sample IC8 represented a permafrost soil from the late Pleistocene Ice Complex (Omolon River), borehole DH-2/07, depth of 16 m. Total carbon concentration was ~1.1% (w/w). Methane levels in all samples tested from this borehole were < 0.0002 mmol kg⁻¹ (Fig. 1B). The age of this sample was estimated to be ~32,000 years, based on the age determination for the same outcrop which was described recently (Legendre et al 2014).

2.1 DNA extraction and sequencing

In the laboratory, material from the inner part of the permafrost cores was subsampled aseptically for DNA isolation. The gDNA was extracted from eight replicates of ~0.5 g each randomly taken from ~50 g of permafrost core collected at corresponding depth (Fig. 1) using the PowerSoil[®] DNA Extraction Kit (MO BIO Laboratories, Inc., USA). Due to low yield (3-6 ng μl^{-1}), gDNAs from eight replicates were combined, then purified and concentrated using Genomic DNA Clean & Concentrator[®] Kit (Zymo Research Corporation, USA).

The gDNA sequencing libraries were prepared using NEBNext[®] reagents (New England BioLabs Inc., USA), according to protocol recommended by the manufacturer, having an estimated peak insert size of 150 nt. Metagenome sequencing was performed at the CRG Genomics Core Facility (Centre for Genomic Regulation, Barcelona, Spain) on an Illumina

HiSeq 2000™ machine using Flow Cell v3 with TruSeq SBS v3 reagents and a 2 x 100 cycle sequencing protocol.

2.2 MG-RAST Analysis

Raw sequencing data, i.e., 19.8 Gb representing 143.7 M sequences with an average length of 138 bp for IC4 and 19.7 Gb representing 131.7 M sequences with an average length of 150 bp for IC8, were uploaded to the MG-RAST server (Meyer et al 2008) for gene calling and annotation under ID 4606864.3 for IC4 and 4606865.3 for IC8. A total of 6.6% (IC4) and 3.4% (IC8) sequences failed to pass the quality control (QC) pipeline, whereas 0.3% of total sequences in both data sets were assigned to ribosomal RNA genes.

For functional assignment, protein sequences of putative ORF were searched against the M5NR non-redundant protein database (Wilke et al 2012) with an e-value threshold of $1e^{-5}$, minimum percentage identity of 60%, and minimum alignment length of 15 aa. The taxonomic assignments of Illumina reads were performed against M5NR and M5RNA databases at default parameters. The best-hit classification method was used in both cases for match assessment.

2.3 Comparative metagenomic data analyses

Statistical analyses of the metagenomes were performed in order to compare community composition and functional profiles of the permafrost samples. Statistical significance was calculated using two-sided Fisher's exact test. The differences between proportions were analyzed in STAMP (Parks and Beiko 2010) using MG-RAST taxonomic profile and the Newcombe-Wilson method (Newcombe 1998) at a 95% confidence interval and with

Storey's FDR correction (Storey and Tibshirani 2003, Storey et al 2004). Original data sets were deposited at the NCBI Sequence Read Archive (SRA) under the accession numbers SRX763249, SRX751044 (Krivushin et al 2015).

3. Results and Discussion

3.1 Community description

The gDNA yield was higher in IC8 sample, with an average of $0.5 \mu\text{g g}^{-1}$ of wet sediment, in comparison to $0.37 \mu\text{g g}^{-1}$ in the IC4 sample. For gDNA isolation, the PowerSoil[®] DNA Extraction Kit (MO BIO Laboratories, Inc., USA) was selected based on its performance and cost per sample. Previously it has been shown that the gDNA extraction efficiency (based on the measurement of mCherry plasmid copies) for this kit was 98%, whereas efficiency of gDNA extraction from permafrost was estimated at $16 \pm 8\%$ (Vishnivetskaya et al 2014). The theoretical levels of the prokaryotic cell populations for studied samples were calculated from the total gDNA recovered, an average genome length of 4.7 Mb for the soil bacterial/archaeal population estimated from the metagenomics data (Angly et al 2009, Raes et al 2007), and a weight of 4.05 fg (Ellenbroek and Cappenberg 1991) for a genome of this size. Accounting for an average extraction efficiency and including a reduction of the total cell population by the eukaryotic component equal to 25% (Raes et al 2007), the estimated prokaryotic cell populations were $4.3 \times 10^8 \text{ cells g}^{-1}$ for IC4 and $5.9 \times 10^8 \text{ cells g}^{-1}$ for IC8.

For gDNA extractions from permafrost, the PowerSoil[®] DNA Extraction Kit (MO BIO Laboratories, Inc., USA) was shown to be reliable providing similar microbial community structure results with minimal variations among triplicates ($P > 0.05$) (Vishnivetskaya et al

2014). Analyses of metagenomes of the two permafrost samples showed that bacterial genes were dominant and 96.4% and 97.7% of sequences were assigned to the domain Bacteria in the IC4 and IC8 samples, respectively. Archaea were the second dominant domain followed by Eukaryotes, while viruses comprised only 0.06% in IC4 and 0.03% in IC8 samples. Archaea and Eukaryotes were more abundant in the IC4 sample compared to the IC8 sample, i.e., 2.4% versus 1.3% and 1.0 versus 0.9%, respectively. Representation of Bacteria in the metagenomic data was significantly higher than Eukaryotes. A similar pattern has been observed in metagenomic datasets obtained from other soils including cold deserts and tundra (Fierer et al 2012). Bailly with co-authors (Bailly et al 2007) showed that inspite the fact that eukaryotic organisms represented a large portion of the microbial biomass in soil detection of eukaryotes in metagenomes is low because they are physically excluded (by filtration or centrifugation on density gradients) from the biomass before DNA extraction. Nevertheless, during current study it was observed that a number of detected unparalleled genes of Eukaryotic organisms including plants, algae, fungi, and protists was 3.4 and 2.6 times higher in comparison to a number of Bacterial unparalleled genes detected in the IC4 and IC8 samples, respectively. Eukaryotic organisms including fungi have genome size of several orders of magnitude higher than the genome size of a bacterium. Eukaryotic genome sizes vary from, for example, 13.8 Mb for the yeast *Schizosaccharomyces pombe* (with 4,800 protein coding genes) to 69 Mb for the ciliate *Paramecium tetraurelia* (39,600 gene models) (Bailly et al 2007). Both of these free-living unicellular eukaryotes were detected in IC4, while only member of the genus *Schizosaccharomyces* were detected in IC8 at lower abundance. It is likely that a metagenomic library based on gDNA can capture a significant functional diversity inspite of low representation of a eukaryotic microbial community.

The most abundant microorganisms were comprised of nine bacterial and one Archaeal phyla, with Proteobacteria being the most plentiful, followed by Actinobacteria, Firmicutes,

Bacteroidetes, Chloroflexi, Planctomycetes, Euryarchaeota, Acidobacteria, Cyanobacteria, and Verrucomicrobia in that order (Fig. 2). The quantities of all other phyla were less than 1%. Sequences affiliated with Actinobacteria ($P < 1e^{-15}$), Firmicutes ($P < 1e^{-15}$), Chloroflexi ($P < 1e^{-15}$), Cyanobacteria ($P < 1e^{-15}$), Acidobacteria ($P < 1e^{-15}$), Verrucomicrobia ($P < 1e^{-15}$) were dominant within IC8, whereas α -, β -, γ -, and δ - Proteobacteria ($P < 1e^{-15}$), Bacteroidetes ($P < 1e^{-15}$), Planctomycetes ($P < 1e^{-15}$), and Euryarchaeota ($P < 1e^{-15}$) dominated within IC4. The microorganisms that are involved into methane formation and methane metabolism, for example methanogenic archaea (phylum Euryarchaeota) or methanotrophic and methylotrophic bacteria (phylum Proteobacteria), were twice more abundant in more reduced IC4 sample ($E_h = 45$ mV) then in less reduced IC8 ($E_h = 167$ mV). Methanogenic and methanotrophic reactions operate literally at opposite redox conditions of -250 mV and +250 mV, respectively. Estimated value of the redox potential in the permafrost samples does not support prevalence neither methanogenesis nor methane oxidation. The observation that the late Pleistocene samples with higher proportion of sands contain low methane concentrations (Bischoff et al 2013) was not supported during current study (Table 1). Microorganisms without strong association to methane metabolism (e.g. phyla Actinobacteria and Chloroflexi) dominated in IC8.

3.2 Dependence of the microbial community structure on permafrost characteristics as detected from genus taxonomical level

3.2.1 Methanogenic community.

As was shown in previous research (Kraev et al 2013, Legendre et al 2014), methane was not detected in the IC8 sample, but its level in the IC4 sample reached 1-2 mmol kg⁻¹; hence, one would expect to find a greater abundance of methanogenic species in the IC4 sample. Indeed, the phylum Euryarchaeota, which contains methanogens, was twice as prevalent in IC4 sample, and methanogenic Archaea constituted

as much as 1.5% of the microbial communities in IC4 (compared to 0.5% in IC8). The most abundant methane-producing species was *Methanosarcina* (0.32% in IC4 versus 0.14% in IC8), followed by *Methanoregula* (0.28% versus 0.03%), *Methanoculleus* (0.1% versus 0.05%), *Methanosphaerula* (0.1% versus 0.03%), *Methanospirillum* (0.1% versus 0.03%), *Methanosaeta* (0.1% versus 0.03%). Twelve other methanogenic species were detected at <0.1% in IC4 and at an even lower percentage in IC8.

3.2.2 Methanotrophic bacteria. In the vicinity of the environments where methane is produced, methane-oxidizing (methanotrophic) bacteria can be found (Khmelenina et al 2002). Thus, we analyzed the presence of methanotrophic DNA in our data. To date two types of methane-oxidizing bacteria are recognized, Type I methanotrophs belonging to γ -Proteobacteria and type II methanotrophs from α -Proteobacteria. Indeed, the phylum Proteobacteria clearly dominated in the IC4 sample (50.0% versus 26.5% in IC8) with the α -Proteobacteria being more abundant in IC4 (32.7%) in comparison to IC8 (12.3%). On a more refined taxonomical scale the most abundant order of α -Proteobacteria was Rhizobiales (24.2% in IC4 versus 7.7% in IC8), which contains both methanotrophic and nitrogen-fixing bacteria. The α -Proteobacteria Type II methanotrophs (3.32% versus 1.06%) were represented by the genera *Methylocella* (0.53% versus 0.14%), *Methylosinus* (0.24% versus 0.07%), *Methylocystis* (0.22% versus 0.06%), and *Methylobacterium* (2.32% versus 0.8%) in IC4 and IC8. The last genus (*Methylobacterium*) is a facultative methylotroph, however, some species are capable of growth on methane. The class γ -Proteobacteria was the most diverse Proteobacteria class (184 species); nonetheless, it was six-fold less plentiful in comparison to α -Proteobacteria. The γ -Proteobacteria Type I methanotrophs, such as *Methylococcus* (0.13% versus 0.11%), *Methylobacter* (0.12% versus 0.05%), and *Methylophaga* (0.02% versus 0.01%) were again more abundant in the IC4 than in the IC8 sample. In this Siberian permafrost, Type II methanotrophs dominated over Type I

methanotrophs, which is similar to methanotrophic bacteria abundance and diversity in Canadian high Arctic permafrost (Lau et al 2015). Methylophiles, as well as the subset methanotrophs, play an essential role in the carbon cycle. Interestingly, obligate methylophilic bacteria belonging to the β -Proteobacteria, such as *Methylobium* (0.17% versus 0.16%), *Methylobacillus* (0.07% versus 0.08%), *Methylovorus* (0.04% versus 0.04%), and *Methylothermus* (0.06% versus 0.05%), were detected in both the IC4 and IC8 samples at equally low levels.

3.2.3 Bacteria of nitrogen cycle. The nitrogen cycle includes several microbial processes such as N_2 fixation, ammonification, nitrification, and denitrification. Nitrogen-fixing bacteria more abundant in the IC4 were represented by the genera *Bradyrhizobium* (1.85% versus 0.5%), *Sinorhizobium* (1.3% versus 0.6%), *Rhizobium* (0.82% versus 0.4%), *Rhodospirillum* (0.63% versus 0.3%), *Azospira* (0.4% versus 0.12%), the γ -Proteobacteria nitrogen-fixing species. Other species involved in the nitrogen cycle from the *Hyphomicrobium* (2.5% versus 0.23%, capable of denitrification with methanol), *Nitrobacter* (1.9% versus 0.6%, capable of oxidizing nitrite into nitrate) and *Rhodopseudomonas* (3.8% versus 1.2%, capable of carbon dioxide and nitrogen fixation) genera were also more abundant in the IC4 sample. However, ammonia-oxidizing and nitrifying bacteria of the class β -Proteobacteria, such as *Nitrosomonas* (~0.1%), *Nitrosospora* (~0.17%), were detected in both samples at the similar level. Other bacteria involved in the nitrogen cycle are members of the phylum Planctomycetes, many of which conduct anaerobic ammonium oxidation or so-called "anammox" metabolism, a process of ammonia oxidation by nitrite involvement to yield nitrogen gas. Four planctomycetes genera were more abundant in IC4 compared to the IC8 sample with *Planctomyces* (0.63% versus 0.34%) being the most abundant, followed by *Pirellula* (0.61% versus 0.28%), *Blastopirellula* (0.57% versus 0.24%), and *Isosphaera* (0.16% versus 0.13%). Some planctomycetes, e.g., *Pirellula*, are able to live in environments with high inorganic sulfate concentrations (Glockner et al 2003). Nitrogen-fixing cyanobacteria slightly dominated in IC8 (0.82%) in comparison to IC4 (0.71%); however, the proportion of nitrogen-fixing to total amount of cyanobacteria was higher in the

IC4 (55.5%) than in the IC8 (48.6%) sample. Another nitrogen-fixing bacterium dominating in IC8 (2.6%) in comparison to IC4 (0.8%) was the actinobacterium *Frankia*, which is characterized by the ability to engage in a symbiotic relationship with plants, producing nitrogen-fixing root nodules.

3.2.4 Sulfate-reducing bacteria. The δ -Proteobacteria were also more abundant in IC4 (5.8%) than in IC8 (4.5%). Sulfate-reducing bacteria, namely, *Desulfovibrio* (0.49% versus 0.41%), *Desulfatibacillum* (0.22% versus 0.08%), *Desulfococcus* (0.18% versus 0.07%), *Desulfobacterium* (0.14% versus 0.04%), *Desulfomicrobium* (0.11% versus 0.06%), and metal-reducing bacteria of the genus *Geobacter* (1.13% versus 0.82%) were more plentiful in the IC4 sample. Two species from the order Syntrophobacterales were more abundant in IC4; these included the strictly anaerobic, sulfate-reducing, propionate-degrading bacterium *Syntrophobacter* (0.33% versus 0.18% in IC8) and the benzoate-degrading bacterium *Syntrophus* (0.62% versus 0.13%). During growth on certain compounds, both of these organisms are known to form syntrophic associations with methanogens, e.g., *Methanospirillum hungatei*, facilitating methane production (Harmsen et al 1998, Jackson et al 1999). The sulfate-reducing bacterium *Desulfotomaculum* from the phylum Firmicutes was found at similar concentration in both samples (i.e., 0.32-0.33%). However, another strictly anaerobic bacterium *Desulfitobacterium*, which is capable of using a wide variety of electron acceptors, such as nitrate, sulfite, metals, humic acids, and halogenated organic compounds can use H₂ as an electron donor to facilitate sulfate reduction and methanogenesis (Villemur et al 2006) was twice as abundant in the IC4 (0.2%) than in the IC8 sample (0.1%). Sulfate-reducing *Thermodesulfovibrio* species from the Nitrospirae division were found at 0.09% and 0.06% in the IC4 and IC8 samples, respectively.

3.3 Similarities and dissimilarities in the microbial communities based on functional annotation

329 In contrast to the taxonomical assignment, the functional annotation of the metagenomes
330 exhibited a similar structure for the microbial communities. Essential portions of sequencing
331 data ranging from 51.9% in IC4 to 49.1% in IC8 were assigned to unknown proteins, while
332 37.6% (IC4) and 44.1% (IC8) were annotated proteins. A comparison of the metagenomes at
333 the function level using the SEED - the database for bioinformatic research (Overbeek et al
334 2005) genome annotations and profile scatter plot showed that the IC4 and IC8
335 metagenomes possess >83% similarity at the function level (Fig. 1S). Functional analysis of
336 the IC4 and IC8 metagenomes demonstrated that among the annotated protein sequences,
337 the most abundant groups represented housekeeping functions, including carbohydrate
338 metabolism (16.6% and 17.7%), amino acid biosynthesis (13.8% and 14.8%), and vitamin
339 and protein metabolism (12.1% and 13.2%). Approximately 4.1% and 3.7% of annotated
340 reads were related to stress response including a high proportion of oxidative stress- (52%
341 and 48%) and osmotic stress- (13% and 12%) related sequences (Fig. 3). The functional
342 potential documented through the metagenome provides the information on how many
343 different genes may exist in a sample, but it does not give any indication if these genes are
344 expressed in permafrost. If slow metabolic activity at subfreezing temperatures in permafrost
345 exists as shown in the laboratory studies (Rivkina et al 2000) then such a community may
346 develop adaptations directed by sediments' physicochemical properties. Correlations
347 between methane concentrations and both methanogenic community structure and functional
348 profile derived from metagenomics shown in current study suggest that the genomic data
349 could be used as biomarkers for methanogenesis, however additional approaches are needed
350 to provide evidence for metabolic activity *in situ*. One of such methods is a phospholipid
351 analysis. Phospholipid fatty acids (PLFAs) and phospholipid ether lipids (PLELs) are
352 suitable biomarkers for estimation of either viable Bacteria and Eukarya or living Archaea,
353 respectively. For example, Bischoff with co-authors correlated PLEL amounts with methane

concentrations in Holocene sediments suggesting living archaea in those permafrost samples (Bischoff et al 2013). Though the same study showed absence of correlations between PLFA and PLEL markers and methane concentrations in the late Pleistocene sediments concluding presence of inactive methanogenic community or storage of methane that was produced and trapped in the past.

3.3.1 Methane metabolism. Analyses using KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto 2000) (Table 2.) demonstrated that several functional genes directly related to methanogenesis are abundant in the IC4 metagenome (Fig. 4), including genes coding for F420-dependent methylene-H4 MPT reductase (Enzyme Commission number, EC 1.5.99.11), formylmethanofuran dehydrogenases (*fmd*) (EC 1.2.99.5), CoB-CoM heterodisulfide reductases (EC 1.8.98.1), F420-reducing hydrogenases (EC 1.12.98.1), and methylenetetrahydromethanopterin dehydrogenases (*mer*) (EC 1.5.99.9). All of these genes, with exception of F420-dependent methylene-H4 MPT reductase, were also present in the IC8 metagenome though in lower amounts. For example, the gene coding for the enzyme catalyzing the last step of methanogenesis, methyl-coenzyme M reductase (MCR) (EC 2.8.4.1), was represented by 1333 hits in IC4 but only by nine hits in the IC8 metagenome (Table 2).

3.3.2 Nitrogen metabolism. The abundance of genes associated with nitrogen metabolism was higher in the IC4 metagenome than in that of IC8 (2.2% versus 1.7%). The percentage of genes related to nitrogen metabolism detected in permafrost was slightly higher than the relative abundance (1.3%) of genes responsible for nitrogen metabolism reported in metagenomes examined from glacier ice of the Northern Schneeferner (Simon et al 2009). Notably, in the IC8 metagenome the gene *nifH* coding for a nitrogenase (EC 1.18.6.1), which performs nitrogen fixation, was represented only by seven sequences. These genes

378 were associated mainly with Proteobacteria (43.7%) represented by *Bradyrhizobium* sp.,
379 *Azoarcus* sp., *Mesorhizobium loti*, and Actinobacteria (*Frankia* sp., 41%). In the IC4
380 metagenome 90 features corresponding to *nifH* gene were detected, and these nitrogenase
381 sequences were linked predominantly with Proteobacteria (33.5%). Nitrogenase genes
382 related to Firmicutes (primarily *Desulfotobacterium hafniense* and '*Alkaliphilus*
383 *metalliredigens*') and Actinobacteria (*Frankia* species) constituted 17.6% and 15%,
384 respectively. Interestingly, *nifH* genes of cyanobacterial origin (primarily linked to *Nostoc*
385 species) were detected exclusively within the IC4 metagenome (12.9%).

386 Genes connected with the denitrification processes (Table 2), such as nitrate reductase (EC
387 1.7.99.4) and nitrite reductase (EC 1.7.2.1) were found in both metagenomes (Fig. 4). The
388 nitrate reductase (*narG*, EC 1.7.99.4) sequences predominating within the IC4 metagenome
389 came from more diverse phylogenetic groups in contrast to *narG* gene from IC8. By
390 contrast, even though nitrite reductase (*nirS*) genes were significantly overrepresented in
391 IC8, their presence was detected in similar phylogenetic groups in both metagenomes (Fig.
392 5). The sequences related to both nitrite reductase (EC 1.7.1.4) and nitric oxide reductase
393 (EC 1.7.2.5) were found in similar phylogenetic groups in both metagenomes with
394 prevalence in IC4 (Fig. 5).

395 The ammonium oxidation pathway was represented by a few sequences related to
396 hydroxylamine oxidase (EC 1.7.3.6) genes in known nitrifying bacteria such as
397 *Nitrosomonas eutropha*, *Nitrosococcus oceani*, *Nitrospira multiformis* and some others.
398 Genes coding for ammonia monooxygenase (EC 1.14.99.39) were not detected in either
399 metagenome by a search with its EC number. However, a search for ammonia
400 monooxygenase using functional hierarchies such as KEGG orthologs yielded ten hits in
401 IC4 and four hits in IC8, while SubSystems showed presence of 509 hits in IC4 and 324 hits

in IC8. The ammonia monooxygenase sequences were annotated as methane monooxygenase (EC1.14.13.25). It should be noted that the particulate methane monooxygenase and ammonia monooxygenase are related and occur in both methanotrophs and ammonia oxidizers (Holmes et al 1995). These enzymes have wide substrate specificity catalyzing the oxidation of various substrates including ammonia, methane, halogenated hydrocarbons, and aromatic molecules (Arp et al 2002). Overall, the low abundance of ammonia-oxidizers presumably represents the rare utilization of this pathway by permafrost bacteria in studied environments. Similar results were also reported for metagenomes from other cold environments, e.g., high Arctic hypersaline subzero spring (Lay et al 2013) and Arctic snow packs (Larose et al 2013).

3.3.3 Sulfur metabolism. Sequences associated with sulfur metabolism were present in both IC4 and IC8 metagenomes and related to both reduction and oxidation (Table 2, and Fig. 6). Overall genes coding for sulfate reduction were more abundant in the IC8 metagenome including genes for sulfate adenylyltransferase (EC 2.7.7.4), phosphoadenylyl-sulfate reductase (EC 1.8.4.8), and ferredoxin-sulfite reductase (EC 1.8.7.1). Higher abundance of enzymes associated with sulfate reduction (Fig. 4) and higher concentration of SO_4^{2-} (Table 1) as well as absence of methane in IC8 sample may suggest that sulfate-reduction probably dominated methanogenesis or sulfate reduction took part in the anaerobic oxidation of methane at time the sediment were formed. Taxonomic distribution of associated species was similar in both metagenomes with the exception of sulfate adenylyltransferase, which was represented in IC4 by sequences related mainly to Proteobacteria and Actinobacteria. In the IC8 sample this gene was of more diverse phylogenetic origin (Fig. 6). A few sulfur oxidation genes detected were associated with *Renibacterium salmoninarum* and *Gordonia bronchialis* in IC8 and *Mycobacterium* species and *Sinorhizobium meliloti* in IC4.

3.3.4 Stress response. Genes associated with stress response were detected in both of the metagenomes (33683 hits in IC4 metagenome and 28557 in IC8) (Table 2). The three most abundant groups present corresponded to oxidative stress, heat shock, and osmotic stress response genes. Sequences related to oxidative stress originated principally from Proteobacteria, Actinobacteria and Firmicutes (5148 features in IC4 and 3832 features in IC8) and included genes for catalase (EC 1.11.1.6), peroxidase (EC 1.11.1.7), and different superoxide dismutases (EC 1.15.1.1). Their occurrence is presumably explained by increased oxygen solubility at low temperatures and associated increase of reactive oxygen species concentration (Chattopadhyay 2006). Sequences related to osmotic stress were represented by the genes involved in the synthesis and uptake of compatible solutes including choline, betaine, periplasmic glucan, and ectoine. Genes for osmoprotectant ABC transporters were also detected. Choline dehydrogenase (EC 1.1.99.1) (222 features in IC4 and 213 features in IC8 mainly from Proteobacteria and Actinobacteria) and betaine-aldehyde dehydrogenase (EC 1.2.1.8) (166 features in IC4 and 186 features in IC8 from Proteobacteria, Actinobacteria and Firmicutes) were the most abundant enzymes of this class. This emphasizes the importance of betaine osmolyte for the osmoprotection of members in microbial communities from subfreezing environments. The genes encoded the heat shock proteins were mainly represented by the chaperone protein DnaK (816 hits in IC4 and 54 in IC8) and its interacting protein DnaJ (759 hits in IC4 and 67 in IC8). These proteins are among the most plentiful chaperons in the bacterial cell and often prevalent in microorganisms from cold environments (D'Amico et al 2006).

3.4. Implication of metagenomic sequencing data for the analysis of the samples origin and evolution.

The application of biological markers for paleo-reconstructions in various environmental sites has been used occasionally. For example, utilization of lipid analyses for petroleum reservoirs formation and maturation (Seifert and Moldowan 1981); analysis of fossil chironomid assemblages in the Holocene lake-sediment cores (south-central Alaska) for evaluating anthropogenic climatic changes and quantitative paleo-temperature reconstructions (Clegg et al 2010); and analyses of fossil ostracodal assemblages from the Arctic seas for reconstruction of coastline and interpretation of environmental differences in Arctic areas (Stepanova et al 2010). To track the occurrence and distribution of microorganisms in the environment, the gDNA and DNA fragments amplified with PCR can be employed as biological markers. Therefore, we anticipated that integration of the next generation sequencing capabilities with phospholipid analysis and approaches in microbial ecology (such as linking microbial community composition and environmental processes involved in carbon, nitrogen and sulfur cycling) could be used for paleo-reconstructions.

The late Pleistocene Ice Complex or Yedoma deposits are widely distributed across northeast Siberia and their distribution relates to a number of geomorphological settings, for example (1) low-elevation coastal mountains as the main sediment sources for Yedoma; (2) Yedoma areas comprises separate bedrock elevations (100–400 m high); and (3) extended lowland areas dominated by large and numerous thermokarst basins and lakes with Yedoma hills present between them (Schirrmeister et al 2013). A question, which processes formed these deposits has been under dispute in the last several decades. Several hypotheses have been proposed about the origin of the late Pleistocene Ice Complex, including eolian (Tomirdiaro et al 1984, Tomirdiaro and Chernen'k'ii 1987), alluvial (Rozenbaum 1981), and polygenetic (Konishchev and Kolesnikov 1981, Sher et al 1987) formation. Different opinions on the origin of these deposits have been summarized in the recent publications of Schirrmeister and co-authors (Schirrmeister et al 2011, Schirrmeister et al 2013). It was

suggested that the ice rich syngenetic permafrost of the late Pleistocene Ice Complex was developed under a cold-arid climate at less hydromorphic conditions than the lake and lake-alluvial sediments (Schirrmeister et al 2013, Rivkina et al 2006). Cold-arid climate was presumably an important factor for Yedoma formation and Bischoff with co-authors found that Yedoma formed on Kurungnakh Island during cold and dry period from ~32,000 to 16,000 years ago had less abundant community of methanogenic archaea and low amounts of methane trapped within the sediments (Bischoff et al 2013). The Yedoma sediments located on the Yana-Indigirka, and Kolyma lowlands have been characterized by the absence of methane (Rivkina et al 2007, Rivkina and Kraev 2008) and much lower redox-potential and iron (Fe^{2+}) content in comparison to permafrost layers of lake and lake-alluvial origin in the same area (Rivkina et al 2006).

A comparison of the two late Pleistocene permafrost metagenomes from samples of different genesis, lake origin from the floodplain and Yedoma, revealed differences in the composition of the microbial community that reflects the conditions under which these deposits were formed. These data uncovered significant distinctions in microbial community compositions between Yedoma and lake-alluvial sediments. The relatively low abundance of methanogenic archaea, limited presence of enzymes from the carbon, and nitrogen cycles, higher abundance of enzymes from sulfate reduction process, as well as the presence of methanotrophic bacteria could explain the absence of methane in Yedoma deposits and provide evidence that the formation of these sediments took place under much more aerobic conditions. In other words, we may assume that during the late Pleistocene period, nearly 30,000 years ago, different environmental conditions predetermined biogeochemical regimes and composition of microbial communities of the studied samples. Similarly to our study the recent survey of soil taxonomic, phylogenetic, and functional diversity have demonstrated

that metagenomic approaches can be used to build a predictive understanding of variations in microbial diversity and functions across terrestrial biomes (Fierer et al 2012).

3.5. Potential limitation of analysis

Shotgun metagenomic sequencing provides insight into the identity and functional diversity of microbial communities. Although this approach is powerful tool in understanding microbial structure and function, there are a number of potential limitations that may bias conclusions. The limitations and biases, which may be introduced throughout sampling and processing of environmental samples, have been reviewed recently (Hazen et al 2013). Keeping in mind that obtaining a representative sample is the most fundamental step in monitoring microbial communities we selected samples based on their similar age, close total carbon content, and significant differences in methane concentrations. We have expected that minimal biases were introduced during coring, transportation and storage, because the permafrost samples were collected follow the same technique (Shi et al 1997), were transported frozen and stored at -20°C in the Laboratory of Soil Cryology, Pushchino. To obtain average gDNA probes and generate enough gDNA for downstream applications we used eight replicates from each permafrost sample and the PowerSoil[®] DNA Extraction Kit (MO BIO Laboratories, Inc., USA), which showed high gDNA extraction efficiency (Vishnivetskaya et al 2014). To ensure gDNA integrity, the gDNA solutions were properly stored at -20°C in the laboratory and shipped on dry ice to the CRG Genomics Core Facility (Centre for Genomic Regulation, Barcelona, Spain) for sequencing. To decrease biases during metagenome sequencing we did not utilize multiple displacement amplification (MDA) to amplify the total community gDNA from low biomass permafrost samples rather we used multiple replicates to obtain enough gDNA for downstream applications. Keeping in mind that interpretation of metagenomic data is highly dependent upon the depth of

sequencing, accurate annotation, and comprehension of database (Hazen et al, 2013),
bioinformatic analyses were performed in the same fashion for both samples that allowed
accurate comparison of these metagenomes.

4. Conclusions

In the current study, we performed a comparative analysis of the two permafrost samples
isolated from lake sediments and ice complex with utilization of metagenomic sequencing
approach. We have shown that the presence of methane in one sample and its absence in the
second are associated with the prevalence of the certain groups of microorganisms and their
corresponding genes. The possible scenario of the development of both sampled Yedoma
and lake sediments can be proposed on the basis of these findings. Analysis of metagenome
from IC4, a sample of the lake sediments, demonstrated occurrence of many
physicochemical reactions such as denitrification, iron reduction and sulfate reduction,
which could reduce environmental redox potential and ultimately create favorable conditions
for development of methanogenic community and methanogenesis. As reflected from the
composition of IC8 metagenome, the sporadic occurrence of such physicochemical reactions
brought to deficiency of methanogenic activity and lack of biogenic methane in the late
Pleistocene ice complex on Kolyma-Indigirka lowland.

The obtained results demonstrate that the metagenomic analysis of permafrost may give
additional information on the environmental conditions during permafrost sediment
formation.

Author contribution:

E.R., L.P. and F.K. designed research; L.S., M.T., L.P., E.R. performed research; L.S. and
M.T. extracted DNA; F.K. performed DNA sequencing; E.R., L.P., K.K., T.V., A.M.

545 analyzed data; E.R. and F.K. contributed new reagents and analytical tools; E.R., L.P., T.V.
546 wrote the paper.

547

548 **Acknowledgements**

549 This work was supported by grants from the Russian Scientific Fund (14-14-01115) to E.R.;
550 from the National Science Foundation (DEB-1442262) to T.V.; and by HHMI International
551 Early Career Scientist Program (55007424), the EMBO Young Investigator Programme,
552 MINECO (BFU2012-31329 and Sev-2012-0208), AGAUR program (2014 SGR 0974) to
553 F.K. The Russian Scientific Fund (14-14-01115) support was allocated for sample
554 collection, gDNA isolation, and analysis of metagenomic data.

555 **References**

- 556 Angly, F. E., Willner, D., Prieto-Davó, A., Edwards, R. A., Schmieder, R., Vega-
557 Thurber, R., Antonopoulos, D. A., Barott, K., Cottrell, M. T., Desnues, C., Dinsdale,
558 E. A., Furlan, M., Haynes, M., Henn, M. R., Hu, Y., Kirchman, D. L., McDole, T.,
559 McPherson, J. D., Meyer, F., Miller, R. M., Mundt, E., Naviaux, R. K., Rodriguez-
560 Mueller, B., Stevens, R., Wegley, L., Zhang, L., Zhu, B., and Rohwer, F.: The GAAS
561 metagenomic tool and its estimations of viral and microbial average genome size in
562 four major biomes, *PLoS Comput Biol* 5, e1000593,
563 doi:10.1371/journal.pcbi.1000593, 2009.
- 564 Anthony Walter, K. M., Zimov, S. A., Grosse, G., Jones, M. C., Anthony, P. M., Chapin III,
565 F. S., Finlay, J. C., Mack, M. C., Davydov, S., Frenzel, P., and Frolking, S.: A shift of
566 thermokarst lakes from carbon sources to sinks during the Holocene epoch, *Nature*, 511,
567 452–456, 2014.
- 568 Arp DJ, Sayavedra-Soto LA, Hommes NG: Molecular biology and biochemistry of
569 ammonia oxidation by *Nitrosomonas europaea*. *Arch Microbiol* 178, 250-255, 2002.
570
- 571 Bailly J., Fraissinet-Tachet L., Verner M.-C., Debaud J.-C., Lemaire M., Wésolowski-
572 Louvel M. and Marmeisse R.: Soil eukaryotic functional diversity, a
573 metatranscriptomic approach, *The ISME Journal*, 1, 632–642,
574 doi:10.1038/ismej.2007.68, 2007.
575
- 576 Bischoff, J., Mangelsdorf, K., Gattinger, A., Schlöter, M., Kurchatova, A.N.,
577 Herzsuh, U., and Wagner, D.: Response of methanogenic archaea to Late
578 Pleistocene and Holocene climate changes in the Siberian Arctic, *Global*
579 *biogeochemical cycles*, 27, 305–317, doi:10.1029/2011gb004238, 2013.
580
- 581 Brouchkov A., Fukuda M.: Preliminary Measurements on Methane Content in
582 Permafrost, Central Yakutia, and some Experimental Data 21. *Permafrost Periglac*,
583 13, 187–197, 2002.
- 584 Chattopadhyay M.K.: Mechanism of bacterial adaptation to low temperature. *J Biosciences*,
585 31, 157-165, 2006.
- 586 Clegg, B. F., Clarke, G. H., Chipman, M. L., Chou, M., Walker, I. R., Tinner, W., and Hu F.
587 S.: Six millennia of summer temperature variation based on midge analysis of lake
588 sediments from Alaska, *Quaternary Sci Rev*, 29, 3308–3316, 2010.
- 589 D'Amico S, Collins T, Marx JC, Feller G, Gerday C.: Psychrophilic microorganisms:
590 challenges for life. *EMBO Rep*, 7, 385-389, 2006.
- 591 Ellenbroek, F. M. and Cappenberg, T.E.: DNA-synthesis and tritiated-thymidine
592 incorporation by heterotrophic fresh-water bacteria in continuous culture, *Appl. Environ*
593 *Microb*, 57, 1675–1682, 1991.

- 594 Fierer N., Leff J.W., Adams B.J., Nielsen U.N., Bates S.T., Lauber C.L., Owens, S., Gilbert
595 J.A., Caporaso J.G.: Cross-biome metagenomic analyses of soil microbial communities and
596 their functional attributes, *Proc Natl Acad Sci USA*, 109, 21390–21395, 2012.
- 597 Gilichinsky, D. and Rivkina, E.: Permafrost microbiology, in: *Encyclopedia of Geobiology*,
598 edited by: Reitner, J. and Thiel, V., Springer, Dordrecht, the Netherlands, 726–732, 2011.
- 599 Glockner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., and Ludwig, W., Gade,
600 D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R., and
601 Reinhardt, R.: Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1,
602 *Proc Natl Acad Sci USA*, 100, 8298–8303, 2003.
- 603 Graham, D. E., Wallenstein, M. D., Vishnivetskaya, T. A., Waldrop, M. P., Phelps, T. J.,
604 Pfiffner, S. M., Onstott, T. C., Whyte, L. G., Rivkina, E. M., Gilichinsky, D. A., Elias, D.
605 A., Mackelprang, R., VerBerkmoes, N. C., Hettich, R. L., Wagner, D., Wulfschleger, S. D.,
606 and Jansson, J. K.: Microbes in thawing permafrost: the unknown variable in the climate
607 change equation, *ISME J*, 6, 709–712, 2012.
- 608 Harmsen, H. J. M., Van Kuijk, B. L. M., Plugge, C. M., Akkermans, A. D. L., De Vos, W.
609 M., and Stams, A. J. M.: *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-
610 degrading sulfate-reducing bacterium, *Int J Syst Bacteriol*, 48, 1383–1387, 1998.
- 611 Hazen T. C., Rocha A. M. and Techtman S. M.: Advances in monitoring environmental
612 microbes, *Current Opinion in Biotechnology*, 24, 526–533, 2013.
- 613 Holmes, A. J., Costello, A., Lidstrom, M. E., and Murrell, J. C.: Evidence that
614 particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily
615 related, *FEMS Microbiol Lett*, 132, 203–208, 1995.
- 616 Jackson, B. E., Bhupathiraju, V. K., Tanner, R. S., Woese, C. R., and McInerney, M. J.:
617 *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids
618 and benzoate in syntrophic association with hydrogen-using microorganisms, *Arch*
619 *Microbiol*, 171, 107–114, 1999.
- 620 Jansson, J. K. and Tas, N.: The microbial ecology of permafrost, *Nature Rev Microbiol*, 12,
621 414–425, 2014.
- 622 Kanehisa M., Goto S.: KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids*
623 *Res*, 28, 27–30, 2000.
- 624 Khmelenina, V. N., Makutina, V. A., Kalyuzhnaya, M. G., Rivkina, E. M., Gilichinsky, D.
625 A., and Trotsenko, Y.: Discovery of viable methanotrophic bacteria in permafrost sediments
626 of northeast Siberia, *Dokl Biol Sci*, 384, 235–237, 2002.
- 627 Konishchev, V. N. and Kolesnikov, S. F.: Specialities of structure and composition of late
628 Cenozoic deposits in the section of Oyogossky Yar, in: *Problems of Cryolithology*, MGU
629 Publishing, Moscow, 107–117, 1981 (in Russian).
- 630 Kraev, G. N., Schultze, E. D., and Rivkina, E. M.: Cryogenesis as a factor of methane
631 distribution in layers of permafrost, *Dokl Earth Sci*, 451, 882–885, 2013.

- 632 Krivushin, K. V., Shcherbakova, V. A., Petrovskaya, L. E., and Rivkina, E. M.:
 633 *Methanobacterium veterum* sp nov., from ancient Siberian permafrost, *Int. J. Syst. Evol.*
 634 *Micr.*, 60, 455–459, 2010.
- 635 Krivushin, K., Kondrashov, F., Shmakova, L., Tutukina, M., Petrovskaya, L., and
 636 Rivkina, E.: Two metagenomes from Late Pleistocene northeast Siberian permafrost,
 637 *Genome Announc*, 3, e01380-14, doi:10.1128/genomeA.01380-14, 2015.
- 638 Larose, C., Dommergue, A., and Vogel, J. G.: Microbial nitrogen cycling in Arctic
 639 snowpacks, *Environ Res Lett*, 8, 035004, 2013.
- 640 Lau, M. C. Y., Stackhouse, B. T., Layton, A. C., Chauhan, A., Vishnivetskaya, T. A.,
 641 Chourey, K., Ronholm, J., Mykytczuk, N. C. S., Bennett, P. C., Lamarche-Gagnon, G.,
 642 Burton, N., Pollard, W. H., Omelon, C. R., Medvigy, D. M., Hettich, R. L., Pfi ner, S. M.,
 643 Whyte, L. G., and Onstott, T. C.: An active atmospheric methane sink in high Arctic mineral
 644 cryosols, *ISME J.*, 9, 1880–1891, 2015.
- 645 Lay C.Y., Mykytczuk N.C., Yergeau E., Lamarche-Gagnon G., Greer C.W., Whyte L.G.:
 646 Defining the functional potential and active community members of a sediment microbial
 647 community in a high-arctic hypersaline subzero spring, *Appl Environ Microb*, 79, 3637–
 648 3648, 2013.
- 649 Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M.,
 650 Poirot, O., Bertaux, L., Bruley, C., Couté, Y., Rivkina, E., Abergel, C., and Claverie, J-M.:
 651 Thirty thousand-year-old distant relative of giant icosahedral DNA viruses with a
 652 pandoravirus morphology, *Proc Natl Acad Sci USA*, 111, 4274–4279, 2014.
 653
- 654 Liebner, S., Wagner, D.: Abundance, distribution and potential activity of methane
 655 oxidizing bacteria in permafrost soils from the Lena Delta, Siberia, *Environmental*
 656 *Microbiology*, 9, 107-117, 2007.
- 657 Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L.,
 658 Blazewicz, S. J., Rubin, E. M., and Jansson, J. K.: Metagenomic analysis of a permafrost
 659 microbial community reveals a rapid response to thaw, *Nature*, 480, 368–371, 2011.
- 660 Meyer, F., Paarmann, D., D’Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T.,
 661 Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and Edwards, R. A.: The
 662 metagenomics RAST server – a public resource for the automatic phylogenetic and
 663 functional analysis of metagenomes, *BMC Bioinformatics*, 9, 386, doi:10.1186/1471-2105-
 664 9-386, 2008.
- 665 Morozova, D., Wagner D.: Stress response of methanogenic archaea from Siberian
 666 permafrost compared with methanogens from nonpermafrost habitats, *FEMS*
 667 *Microbiol Ecol*, 61, 16–25, 2007.
 668
- 669 Newcombe, R. G.: Two-sided confidence intervals for the single proportion:
 670 comparison of seven methods, *Stat Med*, 17, 857–872, 1998.
- 671 Overbeek, R., , Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H-Yu., Cohoon, M.,
 672 Vale’rie de Cre’cy-Lagard, Diaz N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D.,

673 Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A. Iwata-Reuy, D., Jensen, R., Jamshidi, N.,
674 Krause L., Kuba, M., Larsen, L., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H.,
675 Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C.,
676 Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Yu., Zagnitko, O. and Vonstein, V. : The
677 subsystems approach to genome annotation and its use in the project to annotate 1000
678 genomes. *Nucleic Acids Res*, 33, 5691-5702, 2005.

679 Parks, D. H. and Beiko, R. G.: Identifying biologically relevant differences between
680 metagenomic communities, *Bioinformatics*, 26, 715–721, 2010.

681 Raes, J., Korb, J. O., Lercher, M. J., von Mering, C., and Bork, P.: Prediction of effective
682 genome size in metagenomic samples, *Genome Biol*, 8, R10, doi:10.1186/gb-2007-8-1-r10,
683 2007.

684 Rivkina, E. and Kraev, G.: Permafrost degradation and influx of biogeogases into the
685 atmosphere, Ninth International Conference on Permafrost, Fairbanks, USA, University of
686 Alaska Fairbanks, 29 June–03 July, 1499–1504, 2008.

687 Rivkina, E., Laurinavichius, K., McGrath, J., Tiedje, J., Shcherbakova, V., and Gilichinsky,
688 D.: Microbial life in permafrost, *Adv Space Res*, 33, 1215–1221, 2004.

689 Rivkina, E., Kraev, G., Krivushin, K., Laurinavichius, K., Fyodorov-Davydov, D.,
690 Kholodov, A., Shcherbakova, V., and Gilichinsky, D.: Methane in permafrost of
691 northeastern Arctic, *Earth Cryosphere (Russian)*, 10, 23–41, 2006.

692 Rivkina, E., Shcherbakova, V., Laurinavichius, K., Petrovskaya, L., Krivushin, K., Kraev,
693 G., Pecheritsina, S., and Gilichinsky, D.: Biogeochemistry of methane and methanogenic
694 archaea in permafrost, *FEMS Microbiol Ecol*, 61, 1–15, 2007.

695 Rivkina, E. M., Friedmann, E. I., McKay, C. P. and Gilichinsky, D. A.: Metabolic activity of
696 permafrost bacteria below the freezing point, *Appl Environ Microbiol*, 66, 3230–3233,
697 2000.

698 Rozenbaum, G. E.: Special features of lithogenesis of the alluvial planes in the Eastern
699 Subarctic as related to the problem of the Ice (Yedoma) Complex, in: *Problems of*
700 *Cryolithology*, MSU Press, Moscow, 87–100, 1981 (in Russian).

701 Schirrmeister, L., Kunitsky, V., Grosse, G., Wetterich, S., Meyer, H., Schwamborn, G.,
702 Babi, O., Derevyagin, A., and Siegert, C.: Sedimentary characteristics and origin of the
703 Late Pleistocene Ice Complex on North-East Siberian Arctic coastal lowlands and islands –
704 a review, *Quatern Int*, 241, 3–25, 2011.

705 Schirrmeister, L., Froese, D., Tumskey, V., Grosse, G., and Wetterich, S.: Yedoma: late
706 Pleistocene ice-rich syngenetic permafrost of Beringia, in: *Encyclopedia of Quaternary*
707 *Science*, 2nd edn., edited by: Elias, S., Mock, C., and Murton, J., Elsevier, Amsterdam, 542–
708 552, 2013.

709 Seifert, W. K. and Moldovan, J. M.: Paleoreconstruction by biological markers, *Geochim*
710 *Cosmochim Acta*, 45, 783–794, 1981.

- 711 Shcherbakova, V., Rivkina, E., Pecheritsyna, S., Laurinavichius, K., Suzina, N., and
712 Gilichinsky, D.: *Methanobacterium arcticum* sp. nov., a methanogenic archaeon from
713 Holocene Arctic permafrost, *Int J Syst Evol Micr*, 61, 144–147, 2011.
- 714 Sher, A., Kaplina, T., and Ovander, M.: Unified regional stratigraphic chart for the
715 Quaternary deposits in the Yana-Kolyma Lowland and its mountainous surroundings,
716 Explanatory Note, Decisions of Interdepartmental Stratigraphic Conference on the
717 Quaternary of the Eastern USSR, Magadan, 1982, USSR Academy of Sciences, Far-Eastern
718 Branch, North-EasternComplex Research Institute, Magadan, USSR, 29–69, 1987 (in
719 Russian).
- 720 Shi, T., Reeves, R. H., Gilichinsky, D. A., and Friedmann, E. I.: Characterization of viable
721 bacteria from Siberian permafrost by 16S rDNA sequencing, *Microb Ecol*, 33, 169–179,
722 1997.
- 723 Simon, C., Wiezer, A., Strittmatter, A. W., and Daniel, R.: Phylogenetic diversity and
724 metabolic potential revealed in a glacier ice metagenome, *Appl Environ Microb*, 75, 7519–
725 7526, 2009.
- 726 Stepanova, A. Y., Taldenkova, E. E., and Bauch, H. A.: Arctic quaternary ostracods and
727 their use in paleoreconstructions, *Paleontol J*, 44, 41–48, 2010.
- 728 Storey, J. D. and Tibshirani, R.: Statistical significance for genome wide studies, *Proc Nat*
729 *Acad Sci USA*, 100, 9440–9445, 2003.
- 730 Storey, J. D., Taylor, J. E., and Siegmund, D.: Strong control, conservative point estimation
731 and simultaneous conservative consistency of false discovery rates: a unified approach, *J*
732 *Royal Stat Soc B*, 66, 187–205, 2004.
- 733 Tomirdiario, S. V. and Chernen'k'ii, B. I.: Cryogenic Deposits of East Arctic and Sub Arctic,
734 ANSSSR Far-East-Science Center, Magadan, 1987.
- 735 Tomirdiario, S. V., Arslanov, K. A., Chernen'kii, B. I., Tertychnaya, T. V., and Prokhorova,
736 T. N.: New data on formation of loess-ice sequences in Northern Yakutia and ecological
737 conditionsof mammoth fauna in the Arctic during the late Pleistocene, *Reports Academy of*
738 *Sciences USSR* 278, 1446–1449, 1984 (in Russian).
- 739 Villemur, R., Lanthier, M., Beaudet, R., and Lepine, F.: The *Desulfitobacterium* genus,
740 *FEMS Microbiol Rev*, 30, 706–733, 2006.
- 741 Vishnivetskaya, T. A., Layton, A. C., Lau, M. C. Y., Chauhan, A., Cheng, K. R., Meyers, A.
742 J., Murphy, J. R., Rogers, A.W., Saarunya, G. S., Williams, D. E., Pfi ner, S. M., Biggersta ,
743 J. P.,Stackhouse, B. T., Phelps, T. J., Whyte, L., Sayler G. S., and Onstott, T. C.:
744 Commercial DNA extraction kits impact observed microbial community composition in
745 permafrost samples, *FEMS Microbiol Ecol*, 87, 217–230, 2014.
- 746 Walter, K. M., Edwards, M. E., Grosse, G., Zimov, S. A., and Chapin, F. S.: Thermokarst
747 lakes as a source of atmospheric CH₄ during the last deglaciation, *Science*, 318, 633–636,
748 2007.

- 749 Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E. M., Kyrpides, N., Mavrommatis,
750 K.,and Meyer, F.: The M5nr: a novel non-redundant database containing protein sequences
751 and annotations from multiple sources and associated tools, BMC Bioinformatics, 13,
752 141,doi:10.1186/1471-2105-13-141, 2012.
- 753 Yergeau, E., Hogues, H., Whyte, L. G., and Greer, C. W.: The functional potential of high
754 Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses,
755 ISME J, 4, 1–9, 2010.
- 756 Zimov, S. A., Schuur, E. A. G., and Chapin, F. S.: Permafrost and the global carbon budget,
757 Science, 312, 1612–1613, 2006.

758 Table 1. Physicochemical characteristic of two permafrost soils
 759
 760

| Samples | Grain-size distribution | | | Ice content | Eh | C _{tot} | Water extract (1:5) | | | | | | | | |
|---------|-------------------------|-------|------|-------------|-----|------------------|---------------------|------|-------------------------------|-----------------|-------------------------------|------------------|------------------|----------------|-----------------|
| | Sand | Silt | Clay | | | | Dry residue | pH | HCO ₃ ⁻ | Cl ⁻ | SO ₄ ²⁻ | Ca ²⁺ | Mg ²⁺ | K ⁺ | Na ⁺ |
| | % | | | | | | % | mV | % | % | Meq/100g dry soil | | | | |
| IC4 | 79.44 | 15.84 | 4.72 | 17.5 | 45 | 1.1 | 0.130 | 7.22 | 0.12 | 1.40 | 0.14 | 0.32 | 0.42 | 0.06 | 1.33 |
| IC8 | 40.72 | 52.72 | 6.56 | 27.6 | 167 | 1.2 | 0.135 | 8.16 | 0.57 | 0.32 | 0.48 | 0.75 | 0.35 | 0.02 | 0.56 |

761

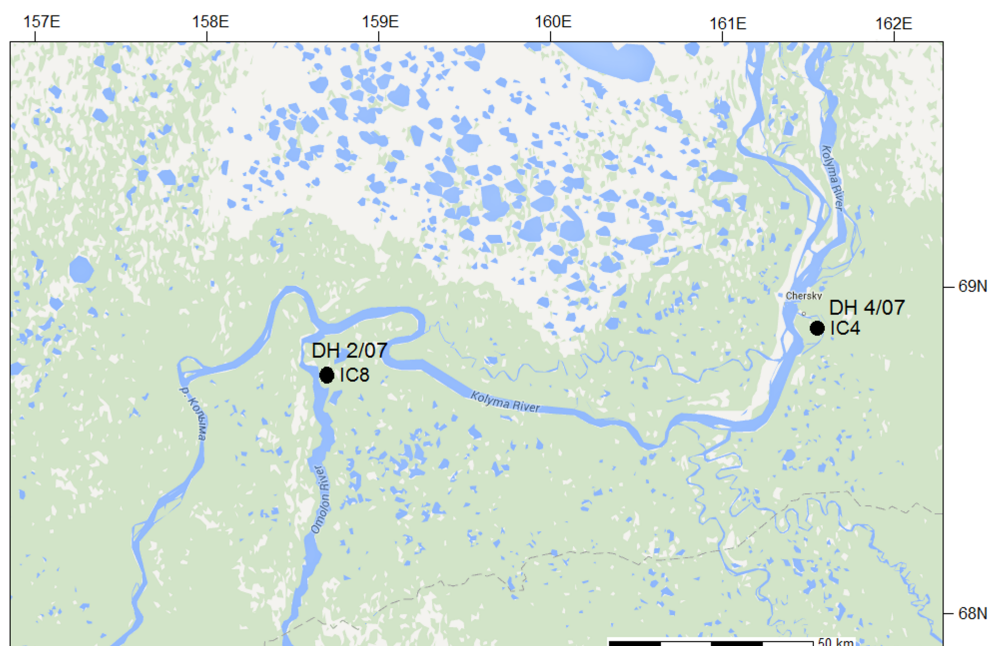
762 **Table 2.** Number of sequences showing homologies to genes associated with KEGG pathways
763 within the IC4 and IC8 metagenomes. Data were generated by searching in Hierarchical
764 classification table based upon SubSystems annotation source.
765

| Enzyme Nomenclature | Function and Enzyme Profile | IC4 | IC8 |
|-------------------------|--|-------|-------|
| | <i>Methane metabolism</i> | | |
| EC 1.5.99.11 | F420-dependent methylene-H ₄ MPT reductase | 3649 | 0 |
| EC 1.12.98.1 | Coenzyme F420 hydrogenase | 1441 | 370 |
| EC 1.8.98.1 | CoB-CoM heterodisulfide reductase | 24601 | 3982 |
| EC 1.2.99.5 | Formylmethanofuran dehydrogenase | 10757 | 3012 |
| EC 1.5.99.9 | Methylenetetrahydromethanopterin dehydrogenases | 370 | 3 |
| EC 2.8.4.1 | Methyl-coenzyme M reductase | 1333 | 9 |
| EC 2.3.1.101 | Formylmethanofuran--tetrahydromethanopterin N-formyltransferase | 3056 | 1642 |
| EC 2.1.1.86 | Tetrahydromethanopterin S-methyltransferase | 1997 | 9 |
| | <i>Nitrogen metabolism</i> | | |
| EC 1.18.6.1 | Nitrogenase (molybdenum-iron) alpha chain nifH | 2519 | 42 |
| EC 1.7.2.1 | Copper-containing nitrite reductase | 1824 | 11405 |
| EC 1.7.2.5 ^a | Nitric-oxide reductase | 3440 | 1838 |
| EC 1.7.1.4 | Nitrite reductase [NAD(P)H] small subunit | 20713 | 13041 |

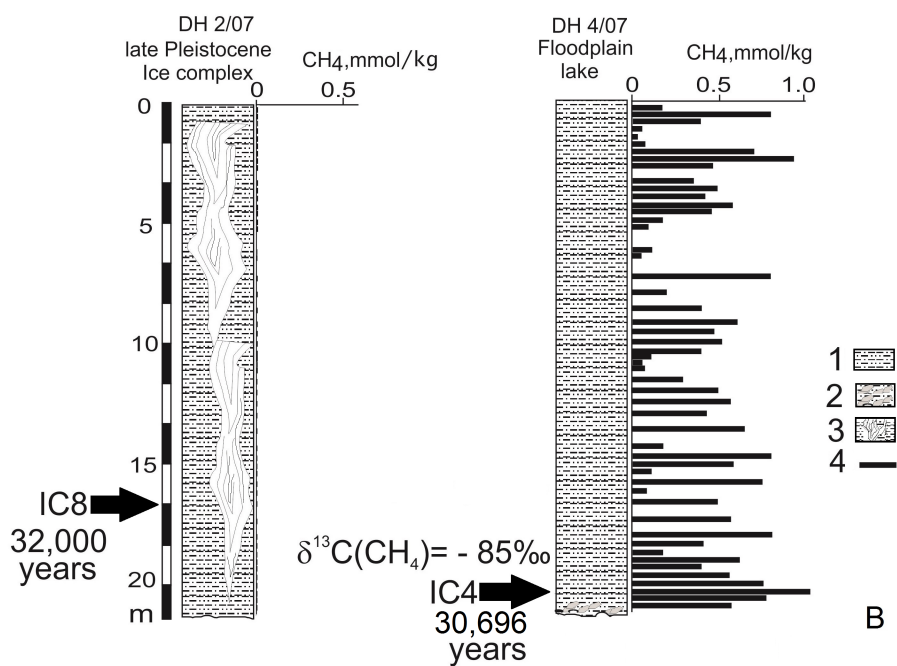
| | | | |
|-------------|--|-------|-------|
| EC 1.7.99.4 | Respiratory nitrate reductase alpha chain narG | 36476 | 28169 |
| EC | Ammonia monooxygenase | 509 | 324 |
| EC 1.7.3.4 | Hydroxylamine oxidase | 517 | 103 |
| EC 1.7.99.6 | Nitrous oxide reductase | 61 | 46 |
| | <i>Sulfur metabolism</i> | | |
| EC 1.8.4.8 | Phosphoadenylyl-sulfate reductase [thioredoxin] | 7521 | 11165 |
| EC 1.8.7.1 | Ferredoxin--sulfite reductase | 2587 | 8920 |
| EC 2.7.7.4 | Sulfate adenylyltransferase subunit 2 | 15449 | 23435 |
| EC 2.7.1.25 | Adenylylsulfate kinase | 22898 | 22418 |
| EC 1.8.3.1 | Sulfite oxidase | 14353 | 13968 |
| | <i>Stress response</i> | | |
| EC 1.11.1.6 | Catalase | 74622 | 41693 |
| EC 1.15.1.1 | Manganese superoxide dismutase | 8922 | 13302 |
| EC 1.11.1.7 | Peroxidase | 35552 | 19698 |
| EC 1.15.1.1 | Superoxide dismutase [Cu-Zn] precursor | 2855 | 1327 |
| EC 1.15.1.1 | Superoxide dismutase [Fe] | 7221 | 3109 |
| EC 1.1.99.1 | Choline dehydrogenase | 25244 | 18018 |
| EC 1.2.1.8 | Betaine-aldehyde dehydrogenase | 10022 | 12782 |
| NR | Chaperone protein DnaK | 78365 | 89944 |
| NR | Interacting protein DnaJ | 32897 | 45343 |

766

767 ^aThe current KEGG number has been created for enzyme EC 1.7.99.7.

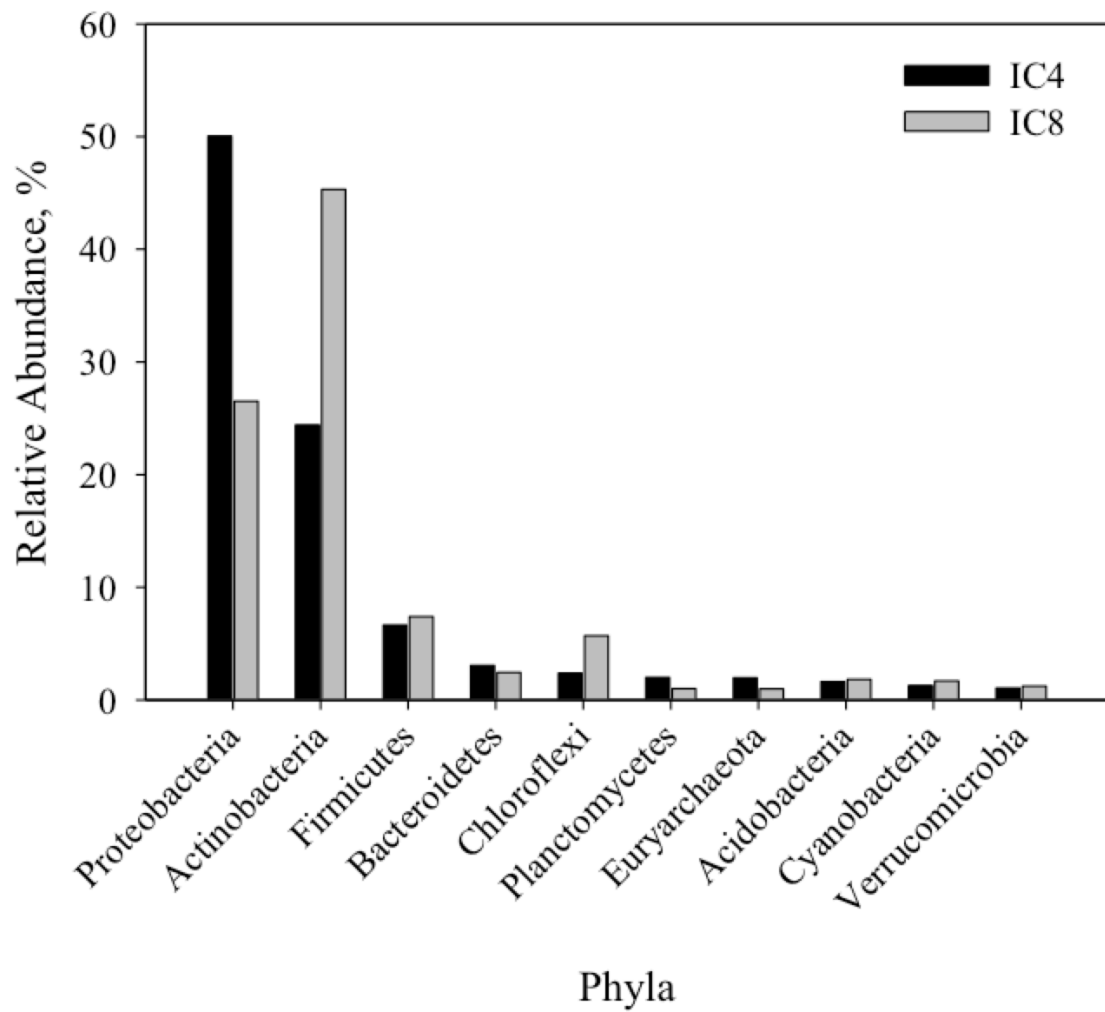


A



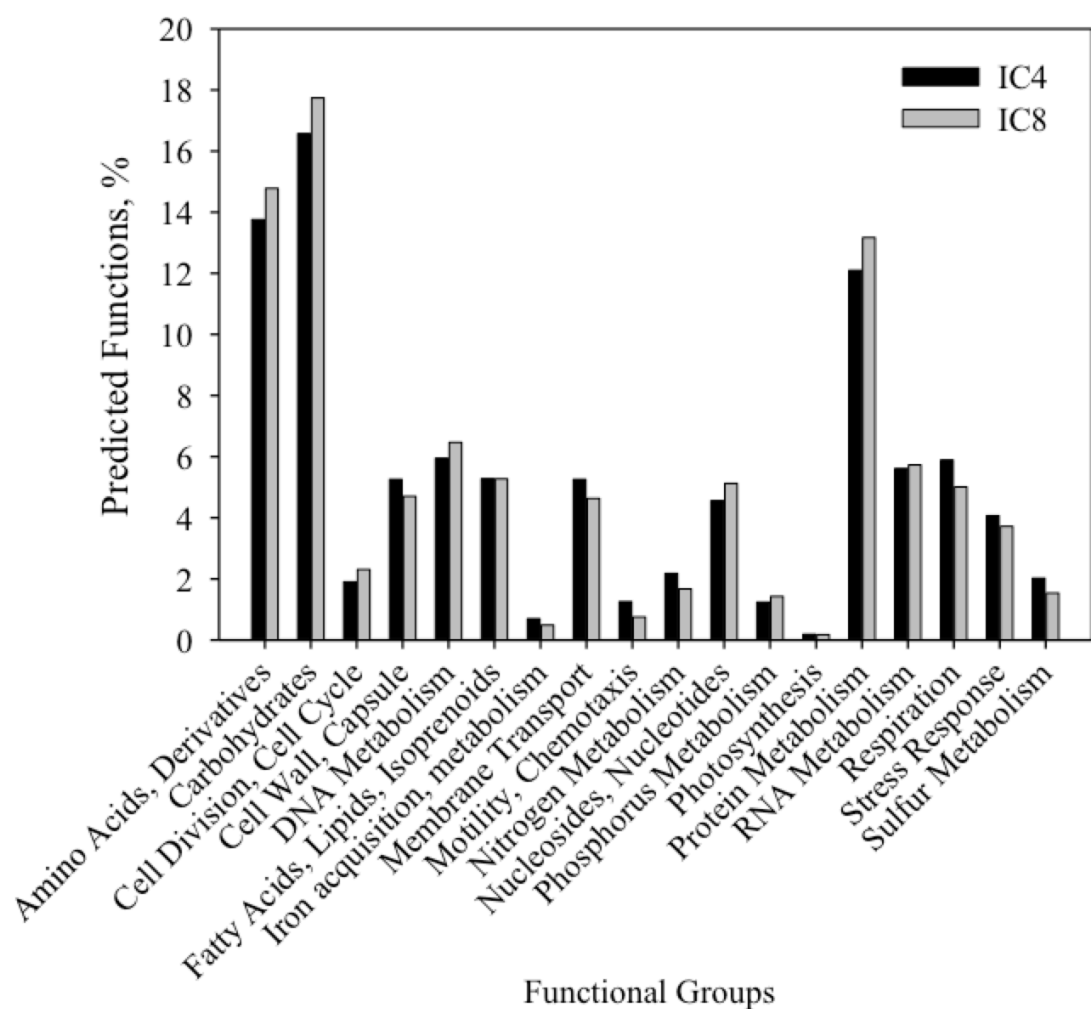
B

Fig. 1. Location of sampling sites on the Kolyma lowland (A). The position of samples IC4 and IC8 in drilling holes and methane content from different depth of these boreholes (B): 1- sandy-loam, 2 - sand with gravel, 3 - ice wedges, 4 - methane content



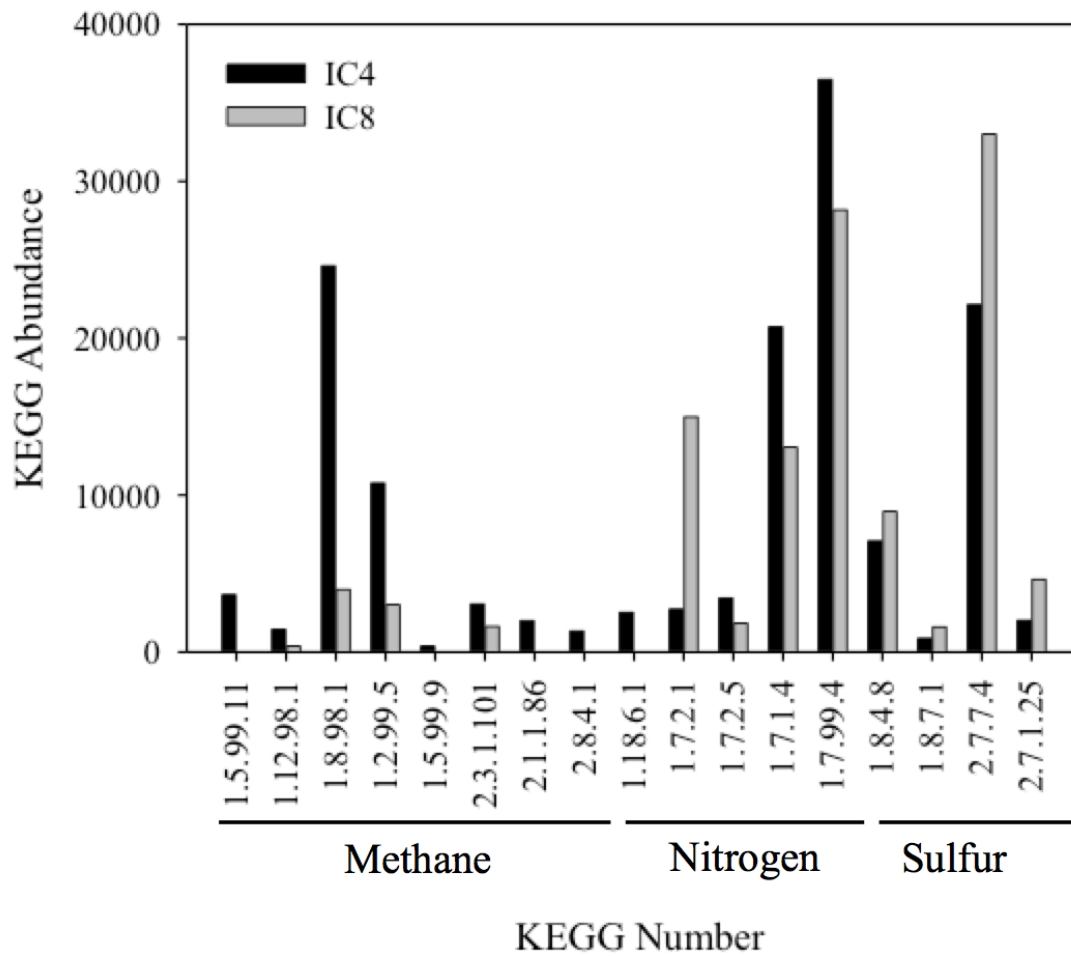
775

776 Fig. 2. Community analyses of the IC4 and IC8 metagenomes at phylum level. The phyla present in
 777 metagenomes at a >1% level are shown.



778

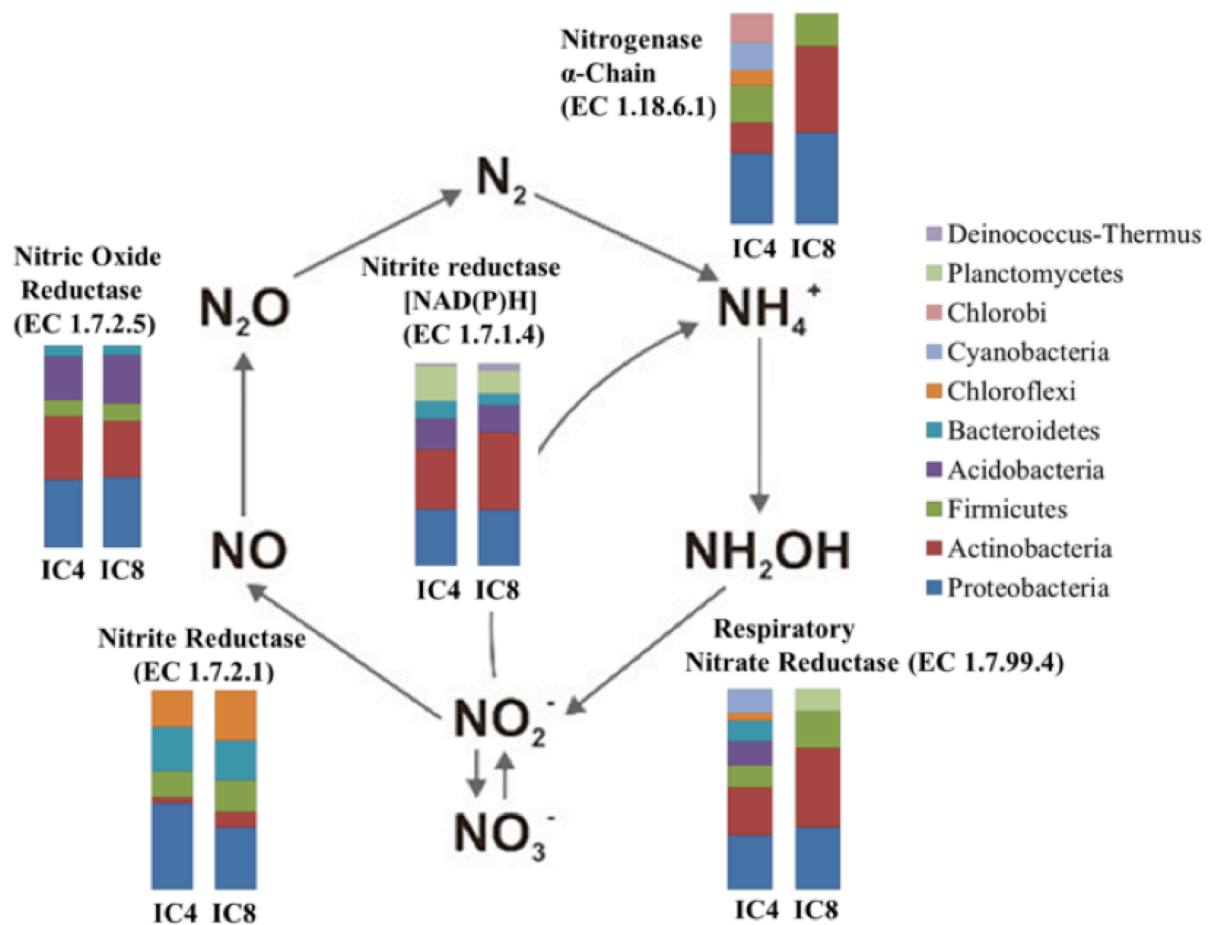
779 Fig. 3. Percentage of gene sequences associated with different functions in the annotated protein
 780 sequences within IC4 and IC8 metagenomes.



781

782 Fig. 4. KEGG abundance of the selected functional gene sequences found within IC4 and IC8
 783 metagenomes. Genes found in low abundance were not included in the figure. Genes from methane,
 784 nitrogen, and sulfur metabolic pathways are underlined. The enzyme name that corresponds to each
 785 KEGG number is given in Table 2.

786



787

788

789 Fig 5. Phylogenetic distribution of the sequences related to nitrogen metabolism within the IC4 and
 790 IC8 metagenomes.

791

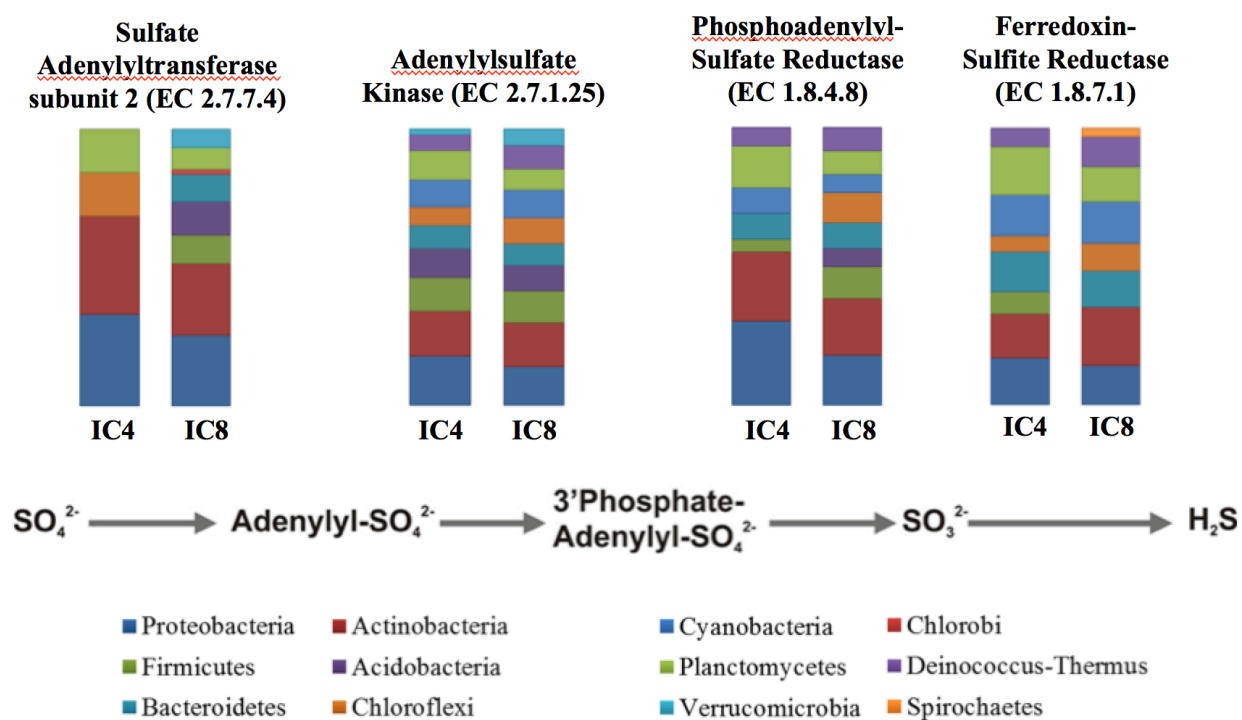


Fig. 6. Phylogenetic distribution of the sequences related to sulfur metabolism (sulfur reduction) within the IC4 and IC8 metagenomes.