ORIGINAL ARTICLE

The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses

Etienne Yergeau[1,2], Hervé Hogues[1], Lyle G Whyte[2] and Charles W Greer[1]

[1]Biotechnology Research Institute, National Research Council of Canada, Montréal, Quebec, Canada and
[2]Department of Natural Resource Sciences, McGill University, Montréal, Quebec, Canada

The fate of the carbon stocked in permafrost following global warming and permafrost thaw is of major concern in view of the potential for increased CH_4 and CO_2 emissions from these soils. Complex carbon compound degradation and greenhouse gas emissions are due to soil microbial communities, but no comprehensive study has yet addressed their composition and functional potential in permafrost. Here, a 2-m deep permafrost sample and its overlying active layer soil were subjected to metagenomic sequencing, quantitative PCR (qPCR) and microarray analyses. The active layer soil and the 2-m permafrost microbial community structures were very similar, with *Actinobacteria* being the dominant phylum. The two samples also possessed a highly similar spectrum of functional genes, especially when compared with other already published metagenomes. Key genes related to methane generation, methane oxidation and organic matter degradation were highly diverse for both samples in the metagenomic libraries and some (for example, *pmoA*) showed relatively high abundance in qPCR assays. Genes related to nitrogen fixation and ammonia oxidation, which could have important roles following climatic change in these nitrogen-limited environments, showed low diversity but high abundance. The 2-m permafrost showed lower abundance and diversity for all the assessed genes and taxa. Experimental biases were also evaluated using qPCR and showed that the whole-community genome amplification technique used caused representational biases in the metagenomic libraries by increasing the abundance of *Bacteroidetes* and decreasing the abundance of *Actinobacteria*. This study describes for the first time the detailed functional potential of permafrost-affected soils.
The ISME Journal (2010) **4**, 1206–1214; doi:10.1038/ismej.2010.41; published online 15 April 2010
Subject Category: microbial ecology and functional diversity of natural habitats
Keywords: arctic soils; carbon cycle; metagenomic sequencing; microarray analysis; nitrogen cycle; permafrost

Introduction

Permafrost underlies $\sim 26\%$ of terrestrial ecosystems (Williams and Smith, 1989) and is estimated to contain around 15% of the world's soil carbon (much of which is frozen as poorly decomposed plant remains) (Post *et al.*, 1982). Recent modeling of permafrost indicates that as much as 90% of the near-surface Arctic permafrost could thaw by 2100 (Lawrence and Slater, 2005), which could lead to an increase in atmospheric CO_2 of roughly 25–50 p.p.m (Ping *et al.*, 2008). The fate of the carbon stocked in permafrost-affected soils will partly depend on the

local environmental conditions, as the relative production of methane and carbon dioxide varies according to soil water/oxygen content. The genetic potential of the microorganisms present in permafrost is also likely to determine whether permafrost environments will become net sources or sinks of greenhouse gases following warming. However, the microbiology of permafrost remains relatively unexplored (reviewed in Steven *et al.* 2006, 2009; Wagner, 2008). The concerns about potential large-scale positive feedback loops on global warming make it crucial to gain in-depth knowledge of the microbes inhabiting permafrost environments.

Processes that will have direct influences on greenhouse gas emissions, such as methane production, methane oxidation and organic matter decomposition, are all mainly carried out by soil microorganisms. The production of CH_4 is solely driven by a small group of strictly anaerobic organisms belonging to the *Euryarchaeota*, which were proven

to be viable inhabitants of permafrost (Wagner et al., 2005; Ganzert et al., 2007; Rivkina et al., 2007). The biological oxidation of methane by methane-oxidizing bacteria (methanotrophs) was shown to be a major sink for methane in the permafrost habitats (Trotsenko and Khmelenina, 2005). However, most of the previous microbiological studies of permafrost focused on 16S rRNA genes, on a few functional genes, on the measurement of general processes or on the isolation and discovery of new microbial species (reviewed in Steven et al. 2006, 2009; Wagner, 2008). Highly parallelized methods such as next-generation sequencing can target all functional genes present in a given sample at once, providing an in-depth assessment of the genetic potential of the microorganisms of an environment.

The main objectives of this study were (1) to describe the genetic potential of indigenous permafrost microorganisms and (2) to identify the functional and phylogenetic differences between permafrost and its overlying active layer soil. To attain these objectives, we sequenced the metagenomes of two samples from a single core from Ellesmere Island in the Canadian high Arctic (core Eur3, Steven et al., 2008) using 454 GS FLX Titanium pyrosequencing. We complemented these results with quantitative PCR (qPCR) analyses targeting a range of genes from different microbes and with microarray analyses of the 16S rRNA genes of microorganisms frequently found in cold environments.

Materials and methods

Permafrost samples
The core (Eur3, Steven et al., 2008) consisted in a continuous core from the surface soil up to 14-m deep permafrost from a single-drill hole, which was collected in May 2003 near Eureka, in the Canadian high Arctic (80°0.029 N, 85°50.367 W). Coring equipment, control of microbiological contamination during coring and subsampling was as previously described (Juck et al., 2005). Core samples were kept frozen during transport and stored at −20 °C until further analysis. The core was sectioned at 1-m depth intervals, and the surface (active layer) and 2-m deep (2-m permafrost) subsections were used in this study. Soil subsamples for molecular analyses were collected under sterile conditions from the center of the core. The active layer soil and the permafrost had the following characteristics (active and 2-m permafrost): pH: 7.9 and 7.8; salinity: 1.5 and 3.1 mEq per 100 g soil dry weight; gravimeteric moisture content: 12.7% and 27.0%; total C: 3% and 1.4%; total N: 0.3% and 0.1% (Steven et al., 2008). The depth of the summer thaw in Eureka is 50–70 cm and the estimated age of the permafrost is 5000–6000 years.

DNA extraction, amplification and 454 sequencing
DNA was extracted from a 10 g subsample using the MoBio PowerMax Soil DNA Isolation kit (MoBio Laboratories, Carlsbad, CA, USA). From the DNA quantities extracted from 10 g of the material, it was calculated that to gather enough DNA for direct 454 shotgun library preparation (that is, 4–5 µg DNA), it would have required >1 kg of the material for the 2-m permafrost sample. This amount of material was not available, hence DNA amplification was necessary. Extracted DNA was then subjected to multiple displacement amplification (MDA) using the illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ, USA) following the manufacturer's instructions for chemical denaturation, except that volumes were increased to 50 µl and amplification was carried out for 4 h. Several amplification reactions were pooled to obtain sufficient DNA for sequencing. Amplified DNA was purified using the QiaEx II Gel Extraction kit (Qiagen, Valencia, CA, USA). Roche 454 GS FLX Titanium sequencing (454 Life Sciences, Branford, CT, USA) was performed at The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada.

Bioinformatic analyses
Replicated sequences were first removed from the data set using the method of Gomez-Alvarez et al. (2009). After this first cleanup step, sequences were assembled in contigs using the Phrap software (http://www.phrap.org/). All contigs and singletons were then annotated using the metagenomics (MG)-RAST server (Meyer et al., 2008). For comparison purposes, the raw (unfiltered and unassembled) sequence data sets were also annotated in MG-RAST. The maximum e-value for a significant match was set to 10^{-5} and the minimum alignment length was set to 50 bp.

PMA treatments
Propidium monoazide (PMA) intercalates into DNA and can be covalently cross-linked to DNA with light exposure, which inhibits subsequent enzymatic reactions. Intact cell membranes are impermeable to PMA and this molecule can thus be used to differentiate naked DNA and dead cells from bacteria with an intact cell membrane (Nocker et al., 2007). To test whether the DNA extracted from the 2-m permafrost and the active layer soil was from viable cells, the samples were treated with PMA following the concentrations and conditions given for soils in previous studies (Pisz et al., 2007; Bae and Wuertz, 2009). The material (10 g) was mixed with 10 ml sterile ultrapure water and 50 µl of 20 mM PMA. This mixture was incubated in the dark for 5 min, shaking manually every minute. The samples were then placed on ice and exposed for 5 min to a 500 W halogen lamp placed 20 cm from

the tubes, shaking manually every minute. No-PMA controls were subjected to the same procedure but without adding PMA. DNA was then extracted directly from the samples as described above. The effect of PMA treatments was evaluated by qPCR analyses at the phylum or class level, as described below.

Real-time qPCR

qPCR was performed on unamplified DNA samples (unless stated otherwise) in 25 μl volumes using the iQ SYBR green Supermix (Bio-Rad Laboratories, Hercules, CA, USA) on a Rotor-Gene 3000 apparatus (Corbett Life Science, Sydney, NSW, Australia) as previously described (Yergeau et al., 2009a). The qPCR conditions and primers are given in Supplementary Table S1. Standards were made from 10-fold dilutions of linearized plasmids containing the gene fragment of interest that was cloned from amplified soil DNA. Phage λ DNA was used to evaluate the presence of PCR inhibitors as previously described (Yergeau et al., 2009a). The inhibition was comparably low for both DNA extracts (active: 1.50%, 2-m permafrost: 2.50%). qPCR values for all other assays were corrected accordingly.

16S rRNA genes microarray

A microarray platform targeting the 16S rRNA genes of Bacteria and Archaea found in cold environments was used as previously described (Yergeau et al., 2009a). To summarize, the 16S rRNA genes were amplified from the raw DNA extracts, chemically labeled and hybridized overnight onto the microarray. Details about the microarray and probe design are available at the National Center for Biotechnology Information (NCBI) GEO database under the platform accession number GPL8953. The presence–absence of the different taxa targeted by the microarray was then used as a community profile to evaluate the similarity between the active layer soil and 2-m permafrost, and to contrast it with previously assessed pristine Arctic surface soil samples from the same region (Yergeau et al., 2009a).

Statistical analyses

All statistical analyses were performed in R (version 2.9.0, The R Foundation for Statistical Computing). Distance matrices were calculated using the 'veg-dist' function of the 'vegan' library and principal coordinate analyses were carried out using the 'cmdscale' function based on the square root of 1-Jaccard (16S rRNA gene microarray) or on the square root of Bray-Curtis (MG-RAST subsystems). For the MG-RAST subsystems, the sums of the relative abundance of selected level 2 subsystems were added to the ordination as supplementary variables that were not involved in the calculation of the ordination.

Data deposition

The microarray data discussed in this publication have been deposited in NCBI GEO (Edgar et al., 2002) and are accessible through GEO series accession number GSE20073 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20073). The sequence data generated in this study was deposited in the Sequence Read Archive of NCBI and are accessible through accession number SRA010889.6.

Results

Metagenomic sequence statistics

Sequencing resulted in 1 103 48 and 1 024 347 sequences for the active layer soil and 2-m permafrost, respectively (Supplementary Table S2). Relatively large proportions (10.3–65.6%) of these sequences were replicates, which were removed from the data set, as they arise from the attachment of DNA to empty beads during emulsion PCR and were not derived independently from the community DNA (Gomez-Alvarez et al., 2009). Only the longest representative of each replicate group was kept for downstream analyses. For the active layer soil, assembly resulted in 125 722 contigs and 276 150 singletons for a total of 401 872 sequences. For the 2-m permafrost, assembly resulted in 8601 contigs and 60 834 singletons for a total of 69 435 sequences. These data sets (contigs + singletons) were used in most downstream analyses.

Phylogenetic community composition

There were significant matches from the active layer (50.76% of all sequences) and the 2-m permafrost (53.50% of all sequences) to organisms present in the MG-RAST database (based on all genes). These sequences mapped to 929 (active layer) and 761 (2-m permafrost) different organisms. Among these, 711 organisms were detected in both samples with the remaining 268 organisms being unique to one sample or the other. Owing to biases related to MDA (see below), the community composition was not calculated from the number of sequences present in the metagenomic data sets. The 16S rRNA gene sequences were not used for community composition assessment primarily not only for the same reason but also because of the low number of 16S rRNA genes recovered (86 bacterial genes for the active layer and 11 bacterial sequences for the 2-m permafrost, with no Archaea sequences for both layers). To have an idea of the phylogenetic community composition in the permafrost samples, a range of phyla or classes was assessed using qPCR. According to qPCR analyses, bacteria dominated the microbial communities of the active layer soil and of the 2-m permafrost (Table 1). Archaea and

Table 1 Real-time PCR quantification results for the active layer soil and the 2-m permafrost

	Active	*2 m*
Ribosomal genes		
Domain		
Archaea	1.49×10^4	1.51×10^2
Total bacteria	3.05×10^7	1.55×10^6
Fungi	8.60×10^4	8.80×10^3
Phylum or classes		
Acidobacteria	6.82×10^5	1.71×10^4
Actinobacteria	7.53×10^6	1.83×10^5
Alpha-	1.08×10^6	3.11×10^4
Bacteroidetes	2.23×10^6	5.34×10^4
Beta-	3.03×10^4	1.92×10^5
Firmicutes	2.54×10^4	5.50×10^3
Methanotrophs		
Type I	1.61×10^5	7.20×10^3
Type II	2.23×10^4	4.87×10^3
Functional genes		
Carbon		
chiA	See text.	See text.
mcrA	7.57×10^2	7.39×10^1
pmoA	4.32×10^5	1.93×10^4
Nitrogen		
Archaea *amoA*	2.05×10^5	2.97×10^4
Bacterial *amoA*	6.54×10^3	5.85×10^3
nifH	1.40×10^5	8.44×10^3

Abbreviations: *Alpha-*, *Alphaproteobacteria*; *Beta-*, *Betaproteobacteria*.

Fungi were detected in much lower abundance (~ 200–$10\,000$ times less; Table 1). Among the bacterial phyla or classes measured, *Actinobacteria* was dominant in both samples (~ 10–25% of the total community), with a codominance of *Betaproteobacteria* ($\sim 15\%$) in the 2-m permafrost (Figure 1a and Table 1). The next most abundant groups were *Bacteroidetes* and *Alphaproteobacteria* ($\sim 5\%$), whereas *Acidobacteria* and *Firmicutes* were the least abundant groups measured (~ 1–2%) (Figure 1a and Table 1).

A microarray targeting the 16S rRNA genes of microorganisms frequently found in cold environments was used as a 'cold-adapted' community fingerprinting method. Even though some genera were retrieved using both methods, the microarray platform is completely independent of the metagenomic data set and was used to confirm the similarity in community structure observed between the two samples. The hybridization patterns of other pristine surface soil samples from the same location (Eureka) and from another location on Ellesmere Island (Alert) from a previous study (Yergeau *et al.*, 2009a) were used here as a comparison for the active layer soil and 2-m permafrost. The active layer DNA hybridized to 33 probes and the 2-m permafrost DNA hybridized to 23 probes. Of these probes, 12 were scored as present in the two samples. The hybridization patterns were used to construct a principal coordinate analysis ordination in which the second axis separated samples from Eureka (including the active layer soil and 2-m permafrost)
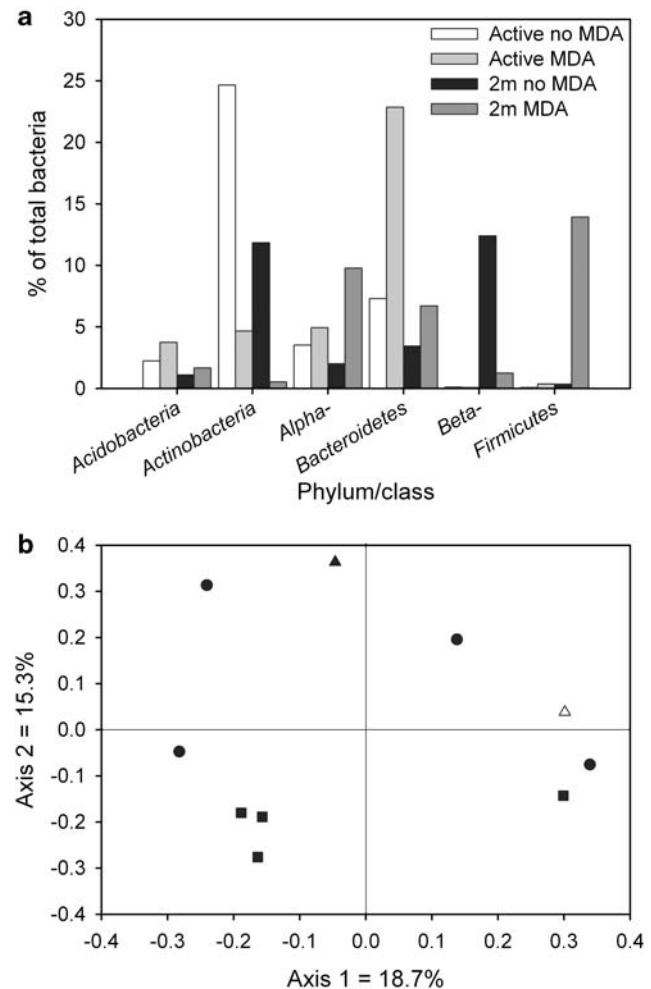

Figure 1 Phylogenetic community composition of the active layer soil and 2-m permafrost. (**a**) Percentage of total bacterial abundance for different phyla or classes based on qPCR quantification for MDA-amplified and unamplified DNA extracted from the active layer soil or 2-m permafrost. Alpha-: *Alphaproteobacteria*, Beta-: *Betaproteobacteria*. (**b**) Principal coordinate analysis based on Jaccard similarity calculated from hybridization patterns of DNA extracted from the active layer soil (▲), the 2-m permafrost (△), four Eureka surface soils (●) and from four Alert surface soils (■) using a 16S rRNA gene microarray targeting microorganisms found in cold environments. The hybridization patterns of the Eureka and Alert soils are from a previous study (Yergeau *et al.*, 2009a).

from the Alert soils (Figure 1b). Among the Eureka soils, the active layer soil and 2-m permafrost did not cluster together and were distributed among the other soils from that location (Figure 1b).

Functional community composition
On the basis of the relative abundance of the different subsystems (sets of related functional genes) represented in MG-RAST, we created an ordination of the active layer and permafrost samples (both the assembled-filtered and the raw sequence data sets) together with several other environmental metagenomic data sets publicly available in MG-RAST (Figure 2). In the ordination, the permafrost samples
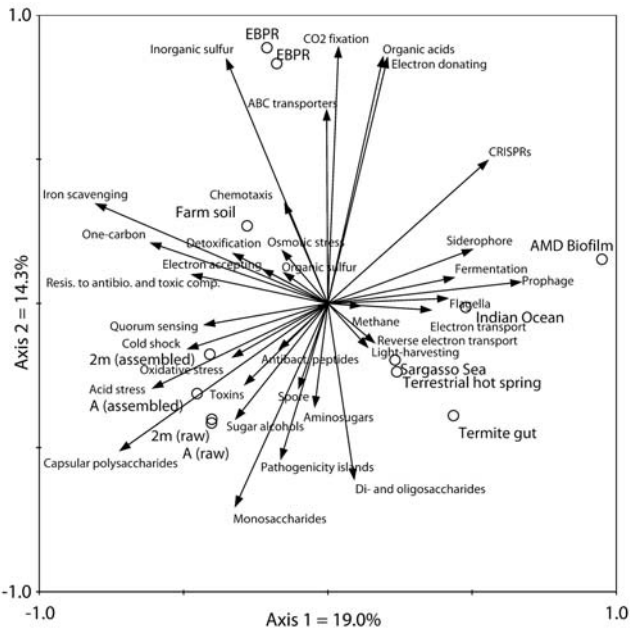
Figure 2 Functional community composition of the active layer soil and the 2-m permafrost. Principal coordinates analysis based on the relative abundance of all MG-RAST subsystems for assembled and raw active layer soil and 2-m permafrost as compared to other publicly available metagenomes. The arrows represent the relative abundance of level 2 subsystems and were added as supplementary variables after the analysis. EBPR, enhanced biological phosphorus removal bioreactors; AMD, acid mine drainage; A, active layer soil; 2 m, 2-m permafrost.

Table 2 Number of different gene variants retrieved in the metagenomic data sets for different functions in the active layer soil and the 2-m permafrost

Function	Active	2 m
Methane		
F420-dependent methylene-H₄ MPT reductase (EC 1.5.99.11)	7	1
Formylmethanofuran dehydrogenase (EC 1.2.99.5)	2	1
Formylmethanofuran-H₄ MPT N-formyltransferase (EC 2.3.1.101)	2	0
Methyl-H₄ MPT:CoM methyltransferase (EC 2.1.1.86)	2	0
Carbohydrates		
α-Amylase (EC 3.2.1.1)	35	14
α-L-fucosidase (EC 3.2.1.51)	23	13
α-N-acetylglucosaminidase (EC 3.2.1.50)	6	1
β-Galactosidase (EC 3.2.1.23)	67	35
β-Glucosidase (EC 3.2.1.21)	30	21
Chitinase (EC 3.2.1.14)	34	8
Trehalase (EC 3.2.1.28)	18	2
Nitrogen		
Ammonia monooxygenase	2	0
Copper-containing nitrite reductase (EC 1.7.2.1)	8	1
Cytochrome c nitrite reductase (EC 1.7.2.1)	1	0
Nitric-oxide reductase (EC 1.7.99.7)	9	5
Nitrite reductase (NAD(P)H) (EC 1.7.1.4)	19	6
Nitrogenase (EC 1.18.6.1)	1	1
Nitrous-oxide reductase (EC 1.7.99.6)	7	2
Respiratory nitrate reductase (EC 1.7.99.4)	13	12

Abbreviation: MPT, methanopterin.

all clustered closely together, apart from all other samples. This clustering was principally because of a higher relative abundance of subsystems related to the response to different stresses, to pathogenicity, toxicity and to the degradation of simple sugars (Figure 2). The only other soil sample (Waseca farm soil) included in the analysis was the most closely related to our samples. Other samples were clearly separated from the active layer and permafrost samples on the first or second ordination axis. This analysis revealed that active layer soil and 2-m permafrost had highly similar levels of functional genes as compared with other samples, and that the filtering and assembly procedures did not cause significant shifts in the functional community composition. These similarities were well mirrored in the number of subsystems detected in each of the samples. For instance, 635 and 620 subsystems were detected (after assembly) in the active layer soil and 2-m permafrost, respectively, among which 576 susbsystems were found in both samples (representing ~90% of the detected subsystems). Similarly, in the active layer data set before assembly, 598 subsystems were detected, among which 589 (98.5%) were also found in the assembled data set. For the 2-m permafrost, 613 subsystems were detected in the metagenomic data set before assembly, among which 598 subsystems (97.6%) were also found in the assembled data set.

Among the many genes detected in the metagenomic libraries, we focused our analyses on methane generation and oxidation, organic matter decomposition and nitrogen cycle genes that might have particular importance in permafrost related to climatic changes. Methanogens were detected in both samples. In the active layer, among the 6763 sequences that mapped to Archaea, 2035 (30.1%) were related to the *Methanomicrobia*, 207 (3.1%) to the *Methanobacteria*, 204 (3.0%) to the *Methanococci* and 38 (0.6%) to the *Methanopyri*. For the 2 m permafrost, 633 sequences were related to the Archaea, among which 242 (38.2%) were related to the *Methanomicrobia*, 30 (4.7%) to the *Methanobacteria*, 23 (3.6%) to the *Methanococci* and 2 (0.3%) to the *Methanopyri*. Several genes related to the generation of methane from the different genera mentioned above were detected (Table 2). The central gene in methane generation, methyl-coenzyme M reductase (*mcr*) was not detected in the metagenomic libraries. However, qPCR analyses revealed that this gene was present at low concentrations in both samples (Table 1).

Type I methanotrophs were detected in metagenomic libraries: 269 and 56 genes were related to *Methylococcaceae* (type I) in the active layer and 2-m permafrost samples, respectively. However, no genes related to type II methanotrophs were detected in the metagenomic libraries. Although type II methanotrophs were detected in large

numbers by qPCR, quantification of the 16S rRNA gene of types I and II methanotrophs indicated that type I methanotrophs were dominant in the active layer and 2-m permafrost samples (Table 1). No *pmoA* sequences were detected in the metagenomic data set, but *pmoA*-specific qPCR assays detected this gene in the active layer soil and 2-m permafrost (Table 1).

Several genes related to the degradation of carbon compounds were detected in the permafrost and active layers, including trehalase, chitinase, β-glucosidase and β-galactosidase (Table 2). These genes were from a diverse range of organisms (~10–70 different organisms), indicating that several members of the microbial community can carry out the functions corresponding to these genes. qPCR confirmed the presence of chitinase genes, but because of the lack of a reliable standard, the exact copy number of this gene could not be evaluated. The cycle threshold (C_t) values were of 18.03 and 26.78 cycles for the active layer soil and 2-m permafrost, respectively. As mentioned above, the *Actinobacteria*, known for their capacity to degrade complex organic matter, were detected in high numbers in both samples. As compared to other published metagenomic data sets, active layer and permafrost habitats contained relatively more genes related to the degradation of sugar alcohols, aminosugars, mono-, di- and oligosaccharides (Figure 2).

Almost all the genes related to the different steps of the N-cycle were detected in both samples (Table 2). A single type of nitrogenase (*nifH*), belonging to *Mesorhizobium loti* (*Alphaproteobacteria*, active layer) or to *Anabaena variabilis* (*Cyanobacteria*, 2-m permafrost) was detected in each of the two samples. qPCR quantification of the *nifH* gene confirmed the presence of this gene in the active layer soil and 2-m permafrost (Table 1). Although ammonia monooxygenase, the key gene in ammonia oxidation, was only detected in the active layer metagenomic data set, several recognized ammonia-oxidizing Bacteria and Archaea were detected in both samples: *Nitrosomonas* (*Betaproteobacteria*), *Nitrospira* (*Betaproteobacteria*), *Nitrosococcus* (*Gammaproteobacteria*) and *Nitrosopumilus* (*Creanarcheota*). qPCR analysis of bacterial and archaeal *amoA* genes confirmed the presence of these genes in both samples (Table 1). The archaeal-to-bacterial *amoA* ratio (measured by qPCR) was 31.3 for the active layer and 5.1 for the 2-m permafrost.

Evaluation of biases related to MDA and the presence of naked DNA
The low biomass present in the 2-m permafrost sample made it necessary for 454 pyrosequencing to amplify the extracted DNA using MDA. The representational biases associated with such a procedure were evaluated to assess the quantitative power of

the data set following amplification. MDA-amplified DNA extracts were compared with the unamplified DNA extracts using qPCR with primers targeting the 16S rRNA genes of six different phyla or classes. The resulting quantitative results indicated that MDA amplification caused major biases in the community composition at the phylum or class level. For instance, MDA decreased the proportion of *Actinobacteria*, whereas it increased the proportion of *Bacteroidetes* for both samples (Figure 1a). For the 2-m permafrost sample, MDA amplification also increased the proportion of *Alphaproteobacteria* and *Firmicutes*, whereas it decreased the proportion of *Betaproteobacteria* (Figure 1a). To evaluate what part of the analyzed DNA was extracted from cells with an intact membrane compared with naked DNA, we treated samples with PMA before DNA extraction. Using the same set of qPCR analyses at the phylum or class level, the comparison of PMA-treated samples with untreated samples did not reveal any major differences (Supplementary Figure S1).

Discussion

This study represents the first metagenomic analysis of permafrost samples. We deliberately choose to focus our sequencing effort on a single core represented by two samples, which allowed us to achieve in-depth comparisons between permafrost and its overlying active layer soil. Owing to the low biomass available in these samples, whole-community genome amplification was carried out before sequencing and caused a representational bias in our data set. This representational bias was seen as an acceptable compromise to access the metagenomic information of permafrost, and was partly counterbalanced by complementary analyses that were performed on the raw DNA extracts.

Phylogenetic and functional community composition
On the basis of qPCR quantification of the unamplified DNA, *Actinobacteria* appeared as the dominant phylum in the active layer soil, whereas *Actinobacteria* and *Betaproteobacteria* codominated the 2-m permafrost. This is in very good agreement with the 16S rRNA genes clone libraries that were constructed from the same samples (Steven *et al.*, 2008) and from a neighboring 9-m deep permafrost sample (Steven *et al.*, 2007). These results also generally agree with the community composition reported from other polar soils (Kobabe *et al.* (2004); Liebner *et al.* (2008); Wagner *et al.* (2009); Yergeau *et al.* (2007, 2009b)). Non-spore forming *Actinobacteria* were reported to dominate ancient permafrost potentially because of their metabolic activity at low temperature and the presence of DNA-repair mechanisms (Johnson *et al.*, 2007). Their abundance in the samples assessed here seems to confirm that

Actinobacteria are well adapted to the conditions occurring in permafrost habitats.

Compared with other publicly available metagenomic data sets (downloaded from MG-RAST) or to other surface soils from the same region (Yergeau *et al.*, 2009a), the active layer soil and the 2-m permafrost had highly similar functional and phylogenetic community compositions. It therefore appears that the microorganisms present in the permafrost layer are a representative subset of the microorganisms present in the active layer and that, consequently, the active layer could be a good proxy to study the microbiology of the near-surface permafrost layers of this site. The long-term exposure to low temperature, low water and nutrient availability probably reduced microbial diversity and abundance in the permafrost, also causing some shifts in the community composition. The larger environmental fluctuations occurring in the active layer soil as compared with the permafrost might also have caused some of the differences observed in community composition. For instance the *Actinobacteria*, *Alphaproteobacteria* and *Bacteroidetes* were less abundant in permafrost in comparison with the active layer and, in contrast, some other groups were more abundant (for example, the *Betaproteobacteria* and the *Firmicutes*) in the permafrost. These latter groups might better resist to long-term exposure to subzero temperatures, take better advantage of the limited availability of both nutrients and water, and/or, in the case of the *Firmicutes*, form spores. In summary, the microorganisms inhabiting the 2-m permafrost sample studied here are most likely originating from the overlying active layer, as was previously suggested for Antarctic permafrost (Gilichinsky *et al.*, 2007). Although it would have to be confirmed using a single core, similarities in community composition (that is, dominance by *Actinobacteria* and *Proteobacteria*) that were observed with a neighboring 9-m deep permafrost sample (Steven *et al.*, 2007) indicate that this relationship between the active layer soil and the permafrost might also hold for deeper permafrost layers.

Functional genes and associated communities
The microorganisms that degrade organic matter in soil are normally highly diverse and the permafrost-affected soils studied here were no exception. The dominant phylum, *Actinobacteria*, contains members that are recognized as degraders of cellulose, chitin and other complex C compounds, which gives them a central role in the carbon cycle and in the turnover of organic matter. In general, the active layer soil and permafrost microorganisms observed in our core seemed particularly well adapted to the degradation of soil carbon, especially when compared with other environments. This indicates that permafrost thawing at this site could rapidly lead to CO_2 emissions after the aerobic degradation of soil carbon.

Recognized methanogens and methanotrophs were detected in the active layer soil and permafrost, but they only formed a minor part of the total microbial community. Changes in environmental conditions could however increase the activity and abundance of these groups and lead to CH_4 emissions from this site. Indeed, methanogens are influenced by the quality of soil organic carbon (Wagner *et al.*, 2005; Ganzert *et al.*, 2007), by water regime (Hoj *et al.*, 2006) and temperature (Wagner *et al.*, 2007). In this study, both qPCR and metagenomic analyses showed that type I methanotrophs were more abundant than type II methanotrophs at the two depths of the core analyzed. This is in line with previous studies from the Canadian Arctic, where type II methanotrophs were not detected among the active microbial community (Martineau *et al.*, 2010). Type II methanotrophs are regularly isolated and can be dominant in Siberian permafrost (Liebner and Wagner, 2007), but PLFA-SIP showed that the communities that were active at low *in situ* temperatures were dominated by type I methanotrophs (reviewed by Wagner, 2008).

Previous studies suggested that Arctic terrestrial environments are generally nitrogen limited (Shaver and Chapin, 1980; Martineau *et al.*, 2010). In that context, the microorganisms responsible for nitrogen fixation and nitrification in permafrost might have an important role in the microbial response to global warming. Nitrogen deficiency could indeed limit the decomposition of organic matter and the subsequent release of greenhouse gases, but this limitation could be relieved if nitrogen fixers and ammonia oxidizers increase their activities after warming. Improved soil nitrogen status in Arctic soils could also increase plant growth and C fixation, counterbalancing a part of the soil C emissions (Mack *et al.*, 2004). Both the nitrogen fixation and ammonia oxidation genes exhibited low diversity in the metagenomic libraries, with one N-fixing species and four ammonia-oxidizer species being detected in each of the samples. This indicates a potentially low redundancy in these functions, which could be vulnerable following acute changes in environmental conditions at this site. However, *nifH* and *amoA* genes were relatively abundant in the core sampled following qPCR assays, indicating that organisms harboring these genes are already present in large numbers and could rapidly increase nitrogen content of the core if provided with favorable environmental conditions. Interestingly, archaeal ammonia oxidation genes were dominant ($\sim$10-fold greater), as previously reported in a range of other soils (Leininger *et al.*, 2006). This dominance was lower in the 2-m permafrost as compared with the active layer soil, indicating some differences in the ammonia-oxidizing community between the two samples.

Experimental biases

We used PMA to differentiate naked DNA from intracellular DNA in the 2-m permafrost and the active layer soil. This is a critical issue when looking at permanently cold environments, as naked DNA can survive for up to 400 000 years in permafrost (Willerslev et al., 2004). Our results indicated that the presence of naked DNA did not cause strong biases in the community composition observed. It therefore appears probable that the majority of the DNA that was extracted from the 2-m permafrost and the active layer soil was from viable cells. It should be stressed that PMA treatment does not indicate that the organisms are active, but that they are viable.

Permafrost environments have low microbial biomass and consequently, it is very difficult, if not impossible, to collect large amounts of DNA from permafrost core samples. The representational biases caused by MDA were therefore seen as a tolerable concession to access, for the first time, the genomic information of permafrost. We accordingly adjusted our metagenomic data analysis to be conservative and to focus only on the presence or absence of different genes, that is, the genetic potential contained within the samples. We also performed complementary analyses on the unamplified DNA to confirm and quantify the trends observed in the metagenomic libraries. Recent studies reported that MDA caused relatively little biases for the sequencing of single genomes (Pinard et al., 2006), but indicated that MDA caused biases in community composition when used on mixed communities (Abulencia et al., 2006; Bodelier et al., 2009). qPCR analyses indicated here that the bias in community composition introduced by MDA is quite strong. In fact, it markedly changed the community composition: from an *Actinobacteria*-dominated community in the unamplified samples to a *Bacteroidetes*-dominated community in the amplified samples. It is uncertain whether these biases are specific to permafrost samples or whether they will be present in other types of environmental samples. It has already been suggested that these representational biases could be partly related to the GC content of the amplified genomes (Bodelier et al., 2009), which is in line with the fact that *Actinobacteria* (high $G+C$) were systematically underestimated and *Firmicutes* (low $G+C$) were overestimated following MDA in our study. These results clearly call for more caution in interpreting quantitative results from metagenomic studies that use whole-community genome amplification. More methodological research is needed in this field, as many environmental samples of interest are too small or too precious to yield the large quantities of DNA required for next-generation sequencing.

Acknowledgements

References

Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH et al. (2006). Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* **72**: 3291–3301.

Bae S, Wuertz S. (2009). Discrimination of viable and dead fecal *Bacteroidales* with PMA-qPCR. *Appl Environ Microbiol* **75**: 2940–2944.

Bodelier PLE, Kamst M, Meima-Franke M, Stralis-Pavese N, Bodrossy L. (2009). Whole-community genome amplification (WCGA) leads to compositional bias in methane-oxidizing communities as assessed by *pmoA*-based microarray analyses and QPCR. *Environ Microbiol Rep* **1**: 434–441.

Edgar R, Domrachev M, Lash AE. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucl Acids Res* **30**: 207–210.

Ganzert L, Jurgens G, Munster U, Wagner D. (2007). Methanogenic communities in permafrost-affected soils of the Laptev Sea coast, Siberian Arctic, characterized by 16S rRNA gene fingerprints. *FEMS Microbiol Ecol* **59**: 476–488.

Gilichinsky DA, Wilson GS, Friedmann EI, McKay CP, Sletten RS, Rivkina EM et al. (2007). Microbial populations in Antarctic permafrost: biodiversity, state, age, and implication for astrobiology. *Astrobiology* **7**: 275–311.

Gomez-Alvarez V, Teal TK, Schmidt TM. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J* **3**: 1314–1317.

Hoj L, Rusten M, Haugen LE, Olsen RA, Torsvik VL. (2006). Effects of water regime on archaeal community composition in Arctic soils. *Environ Microbiol* **8**: 984–996.

Johnson SS, Hebsgaard MB, Christensen TR, Mastepanov M, Nielsen R, Munch K et al. (2007). Ancient bacteria show evidence of DNA repair. *Proc Natl Acad Sci USA* **104**: 14401–14405.

Juck DF, Whissell G, Steven B, Pollard W, McKay CP, Greer CW et al. (2005). Utilization of fluorescent microspheres and a green fluorescent protein-marked strain for assessment of microbiological contamination of permafrost and ground ice core samples from the Canadian High Arctic. *Appl Environ Microbiol* **71**: 1035–1041.

Kobabe S, Wagner D, Pfeiffer EM. (2004). Characterisation of microbial community composition of a Siberian tundra soil by fluorescence *in situ* hybridisation. *FEMS Microbiol Ecol* **50**: 13–23.

Lawrence DM, Slater AG. (2005). A projection of severe near-surface permafrost degradation during the 21st century. *Geophys Res Lett* **32**: L24401.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.

Liebner S, Harder J, Wagner D. (2008). Bacterial diversity and community structure in polygonal tundra soils from Samoylov Island, Lena Delta, Siberia. *Int Microbiol* **11**: 195–202.

Liebner S, Wagner D. (2007). Abundance, distribution and potential activity of methane oxidizing bacteria in permafrost soils from the Lena Delta, Siberia. *Environ Microbiol* **9**: 107–117.

Mack MC, Schuur EAG, Bret-Harte MS, Shaver GR, Chapin FS. (2004). Ecosystem carbon storage in Arctic

tundra reduced by long-term nutrient fertilization. *Nature* **431**: 440–443.

Martineau C, Whyte LG, Greer CW. (2010). Stable isotope probing analysis of the diversity and activity of methanotrophic bacteria in soils from the Canadian high Arctic. *Appl Environ Microbiol* Accepted pending revision.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.* (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Nocker A, Sossa-Fernandez P, Burr MD, Camper AK. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. *Appl Environ Microbiol* **73**: 5111–5117.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN *et al.* (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**: 216.

Ping CL, Michaelson GJ, Jorgenson MT, Kimble JM, Epstein H, Romanovsky VE *et al.* (2008). High stocks of soil organic carbon in the North American Arctic region. *Nat Geosci* **1**: 615–619.

Pisz JM, Lawrence JR, Schafer AN, Siciliano SD. (2007). Differentiation of genes extracted from non-viable versus viable micro-organisms in environmental samples using ethidium monoazide bromide. *J Microbiol Meth* **71**: 312–318.

Post WM, Emanuel WR, Zinke PJ, Stangenberger AG. (1982). Soil carbon pools and world life zones. *Nature* **298**: 156–159.

Rivkina E, Shcherbakova V, Laurinavichius K, Petrovskaya L, Krivushin K, Kraev G *et al.* (2007). Biogeochemistry of methane and methanogenic archaea in permafrost. *FEMS Microbiol Ecol* **61**: 1–15.

Shaver GR, Chapin FS. (1980). Response to fertilization by various plant-growth forms in an Alaskan tundra— nutrient accumulation and growth. *Ecology* **61**: 662–675.

Steven B, Briggs G, McKay CP, Pollard WH, Greer CW, Whyte LG. (2007). Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol Ecol* **59**: 513–523.

Steven B, Leveille R, Pollard WH, Whyte LG. (2006). Microbial ecology and biodiversity in permafrost. *Extremophiles* **10**: 259–267.

Steven B, Niederberger TD, Whyte LG. (2009). Bacterial and archaeal diversity in permafrost. In: Margesin R (ed). *Permafrost Soils*. Springer-Verlag: Berlin. pp 59–72.

Steven B, Pollard WH, Greer CW, Whyte LG. (2008). Microbial diversity and activity through a permafrost/ ground ice core profile from the Canadian high Arctic. *Environ Microbiol* **10**: 3388–3403.

Trotsenko YA, Khmelenina VN. (2005). Aerobic methanotrophic bacteria of cold ecosystems. *FEMS Microbiol Ecol* **53**: 15–26.

Wagner D. (2008). Microbial communities and processes in Arctic permafrost environments. In: Dion P, Nautiyal CS (eds). *Microbiology of Extreme Soils*. Springer-Verlag: Berlin. pp 133–154.

Wagner D, Gattinger A, Embacher A, Pfeiffer EM, Schloter M, Lipski A. (2007). Methanogenic activity and biomass in Holocene permafrost deposits of the Lena Delta, Siberian Arctic and its implication for the global methane budget. *Glob Change Biol* **13**: 1089–1099.

Wagner D, Kobabe S, Liebner S. (2009). Bacterial community structure and carbon turnover in permafrost-affected soils of the Lena Delta, northeastern Siberia. *Can J Microbiol* **55**: 73–83.

Wagner D, Lipski A, Embacher A, Gattinger A. (2005). Methane fluxes in permafrost habitats of the Lena Delta: effects of microbial community structure and organic matter quality. *Environ Microbiol* **7**: 1582–1592.

Willerslev E, Hansen AJ, Rønn R, Brand TB, Barnes I, Wiuf C *et al.* (2004). Long-term persistence of bacterial DNA. *Current Biol* **14**: R9–R10.

Williams PJ, Smith MW. (1989). *The Frozen Earth: Fundamentals of Geocryobiology*. Cambridge University Press: Cambridge, 306pp.

Yergeau E, Arbour M, Brousseau R, Juck D, Lawrence JR, Masson L *et al.* (2009a). Microarray and real-time PCR analyses of the responses of high Arctic soil bacteria to hydrocarbon pollution and bioremediation treatments. *Appl Environ Microbiol* **75**: 6258–6267.

Yergeau E, Newsham KK, Pearce DA, Kowalchuk GA. (2007). Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ Microbiol* **9**: 2670–2682.

Yergeau E, Schoondermark-Stolk SA, Brodie EL, Déjean S, DeSantis TZ, Gonçalves O *et al.* (2009b). Environmental microarray analyses of Antarctic soil microbial communities. *ISME J* **3**: 340–351.