ORIGINAL ARTICLE

Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms

Alexander Tveit[1], Rainer Schwacke[1], Mette M Svenning[1] and Tim Urich[2]

[1]*Department of Arctic and Marine Biology, University of Tromsø, Tromsø, Norway and* [2]*Department of Genetics in Ecology, University of Vienna, Vienna, Austria*

A substantial part of the Earths' soil organic carbon (SOC) is stored in Arctic permafrost peatlands, which represent large potential sources for increased emissions of the greenhouse gases CH_4 and CO_2 in a warming climate. The microbial communities and their genetic repertoire involved in the breakdown and mineralisation of SOC in these soils are, however, poorly understood. In this study, we applied a combined metagenomic and metatranscriptomic approach on two Arctic peat soils to investigate the identity and the gene pool of the microbiota driving the SOC degradation in the seasonally thawed active layers. A large and diverse set of genes encoding plant polymer-degrading enzymes was found, comparable to microbiotas from temperate and subtropical soils. This indicates that the metabolic potential for SOC degradation in Arctic peat is not different from that of other climatic zones. The majority of these genes were assigned to three bacterial phyla, Actinobacteria, Verrucomicrobia and Bacteroidetes. Anaerobic metabolic pathways and the fraction of methanogenic archaea increased with peat depth, evident for a gradual transition from aerobic to anaerobic lifestyles. A population of CH_4-oxidising bacteria closely related to *Methylobacter tundripaludum* was the dominating active group of methanotrophs. Based on the in-depth characterisation of the microbes and their genes, we conclude that these Arctic peat soils will turn into CO_2 sources owing to increased active layer depth and prolonged growing season. However, the extent of future CH_4 emissions will critically depend on the response of the methanotrophic bacteria.
The ISME Journal (2013) **7**, 299–311; doi:10.1038/ismej.2012.99; published online 6 September 2012
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: microbial communities; metagenomics; metatranscriptomics; Arctic peat soils; soil organic carbon; methane

Introduction

Recent estimates point out that 277 Pg of soil organic carbon (SOC) are stored in Arctic peatlands (Tarnocai *et al.*, 2009), which corresponds to 1/3 of the CO_2 in the atmosphere. These peat soils have acted as carbon sinks since the Holocene (Post *et al.*, 1982; Harden *et al.*, 1992). In contrast, they are substantial sources of methane (CH_4), releasing ~35 Tg per year, ~6% of the global CH_4 emissions (Cao *et al.*, 1996). CH_4 is a more potent greenhouse gas than carbon dioxide (CO_2), having 25 times the global warming potential of CO_2 on a 100-year timescale (Forster *et al.*, 2007). Arctic and especially high-Arctic regions are already exposed to, and predicted to experience, a strong temperature increase until the end of the century (4–8 °C higher

annual surface air temperatures) (IPCC, 2007), which is expected to lead to expanded frost-free vegetation periods and increased active layer depths in permafrost soils (Tarnocai *et al.*, 2009).

The major factors in SOC degradation in peatlands, like in mineral soils, are microorganisms of the bacterial, archaeal and eukaryotic (fungi) domains of life, participating in a cascade of aerobic and anaerobic degradation steps, eventually resulting in the emission of CH_4 and CO_2. Major parts of the peat carbon are plant polymers such as cellulose and hemicellulose. The degradation of these to oligomeric and monomeric sugars is considered one of the most important steps in anaerobic degradation of SOC (Kotsyurbenko, 2005) and is catalysed by a diverse set of hydrolytic extracellular enzymes produced by microorganisms. Further steps in the anaerobic degradation are carried out through anaerobic respiration (like denitrification), fermentative and methanogenic pathways, while microbial methane oxidation constitutes the biological filter for methane emissions from peat. Phenolic substances accumulate to high concentrations in the anaerobic, water-logged layers of peat soils, partly owing to low activity of phenol oxidases,

300

which require oxygen for function. The inhibiting effect of phenolic substances has been suggested as a major factor for the low SOC degradation rates in peat soils (Fenner and Freeman, 2011).

Several studies have targeted the microorganisms involved in CH_4 formation (for example, Ganzert et al., 2007; Kotsyurbenko et al., 2007; Metje and Frenzel, 2007; Høj et al., 2008) and oxidation (for example, Wartiainen et al., 2003; Liebner et al., 2009; Martineau et al., 2010; Graef et al., 2011) in Arctic soils. Broader studies have also been conducted, for example, on the biogeography of soil bacterial communities in different climatic zones (Chu et al., 2010), the dynamics of Arctic soil microbial communities in relation to the composition of plant communities (Zak and Kling, 2006) and the availability of SOC (Waldrop et al., 2010 and Coolen et al., 2011). In recent years, metagenomics and metatranscriptomics have developed into powerful tools in microbial ecology that enable non-targeted studies of the genetic potential, gene expression and the composition of soil microbial communities (for example, Tringe et al., 2005 and Urich et al., 2008). Until now only two metagenomic studies on Arctic soils have been conducted, comparing the community in the permafrost with its overlaying active layer (Yergeau et al., 2010) and investigating the response of a permafrost microbial community to thaw (Mackelprang et al., 2011).

It is in the active layers, where the SOC degradation in Arctic peat soils occurs, that more knowledge about the processes and organisms involved is required to understand these and to predict the magnitude of CH_4 and CO_2 emissions from Arctic peat in a warmer climate. We have in this study applied, for the first time, a combined metagenomic and metatranscriptomic approach on high-Arctic peat soils to obtain on-site information about the genomic potential for SOC transformation and to identify the active microorganisms driving these processes. We compared the Arctic peat metagenomes with metagenomes from temperate soils and proposed a model of SOC degradation in Arctic peat, including anaerobic respiratory and fermentative pathways, and methanogenic and methanotrophic microorganisms. The data allow some predictions about future CH_4 and CO_2 emissions from Arctic peat soils.

Materials and methods

Study sites and sampling
We have investigated the active layers of two high-Arctic fens on Svalbard (Norway). Solvatn (N78°55.550 E11°56.611) and Knudsenheia (N78°56.544 E11°49.055) were sampled in August 2009 at the peak of the growing season (Supplementary Figure S1 and S2). Both sites are located in the vicinity of the research settlement Ny-Ålesund. Solvatn is situated on a marine terrace right next to the settlement, while Knudsenheia is located ~5 km northwest of Ny-Ålesund. For both sites,

Knudsenheia and Solvatn, two biological replicates were prepared. Three peat blocks ($20 \times 30 \times 20$ cm) were cut from each of the replicate sites, S1 and S2 (Solvatn) and K1 and K2 (Knudsenheia; see Supplementary Figures S1–S3), and transported to the laboratory within 2 h after sampling. The active layer depths was ~40 cm at both sites. The peat blocks were subsequently stored for 2–3 h at 4 °C until further processing. The three blocks were separated into upper (more oxic) and lower (more anoxic) layers, and the corresponding layers, within each replicate site, were pooled to account for heterogeneity of each replicate site (Supplementary Figure S3). The layer separation was based on visual distinction of the layers based on its colour and structure. The lower layers were processed under nitrogen atmosphere to avoid extensive oxygen contamination. After pooling, samples were immediately shock frozen in liquid nitrogen. The whole procedure took ~5–10 min. Soil pH was measured using the KCl extraction method on fresh peat samples. Water content was estimated by drying peat over night at 120 °C and reweighing the samples. To estimate organic matter contents, the dried samples were combusted at 450 °C overnight. The samples were reweighed, and the ash weight was calculated and added to the weight of organic matter. Concentrations of volatile fatty acids and ethanol were determined by high pressure liquid chromatography analysis of pore water (Metje and Frenzel, 2005) extracted from soil samples before freezing. Results were integrated using peak simple version 393, and 1 mM standards.

Nucleic acids extraction, reverse transcription and 454 sequencing
From each layer of the biological duplicates (Supplementary Figure S3), six or more parallel extractions of nucleic acids were performed using a modified version of the Griffith's protocol (Urich et al., 2008). In samples for metatranscriptomic analysis, DNA was digested using the RQ1 DNase treatment (Promega, Madison, WI, USA), followed by RNA purification using the MEGAclear kit (Ambion, Austin, TX, USA). Samples were subjected to reverse transcription using the Superscript II double-strand complementary DNA synthesis kit (Invitrogen, Carlsbad, CA, USA), following the manufacturer's protocol, with the exception that both first- and second-strand synthesis was carried out for 4 h. RNA template addition was in the range of 500–1000 ng. Concentrations of double-stranded complementary DNA were estimated by SYBR Green I (Invitrogen) assay (Leininger et al., 2006). Nucleic acids for metagenomic analysis were subjected to RNase A treatment for 20 min at 37 °C, followed by phenol:chloroform:isoamylalcohol extraction and chloroform:isoamylalcohol extraction. DNA concentrations were estimated using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Madison, WI, USA). Roche 454 GS FLX Titanium sequencing

(454 Life Sciences, Branford, CT, USA) was carried out at the CEES at the University of Oslo.

Bioinformatic analysis
Metatranscriptomic and metagenomic sequences were first filtered using LUCY (Chou and Holmes, 2001), removing short (<150 bp) and low-quality sequences (>0.2% error probability). Metatranscriptomic sequences stemming from ribosomal RNA (rRNA) and putative messenger RNAs (mRNAs) were separated by comparing all sequences against a combined database of small and large subunits of rRNA (SSU and LSU rRNA) using BLASTN (Altschul *et al.*, 1997) and MEGAN (Huson *et al.*, 2007; Urich *et al.*, 2008). Sequences with a bit score <50 were assigned as putative mRNA tags. Metagenomic and putative mRNA tags in the metatranscriptomes were subjected to the 454 Replicate filter (Gomez-Alvarez *et al.*, 2009) for the removal of artificially replicated sequences. The filter was applied to remove exact duplicate sequences from the metatranscriptomes (settings: sequence identity cutoff: 1.0, length difference requirement: 1.0 and number of beginning base pairs to check: 3). For the metagenomes, the default settings were applied.

SSU ribo-tags were taxonomically assigned by MEGAN analysis of a BLASTN file against a SSU rRNA reference database (parameters: min. bit score 150, min. support 1, top percent 10; 50 best blast hits) (Urich *et al.*, 2008; Lanzen *et al.*, 2011). SSU ribo-tags assigned to the order Methylococcales were assembled into ribo-contigs using CAP3 (Huang and Madan, 1999), using two subsequent rounds of assembly with (1) a minimum overlap of 150 bp with a minimum similarity threshold of 99% and mismatch and gap scores of −130 and 150, and (2) minimum overlap of 150 bp and minimum 97% similarity threshold, respectively. Contigs were uploaded in ARB (http://www.arb-home.de/), and contigs longer than 1300 bp together with selected reference sequences were used to generate a maximum likelihood tree with the RAxML (Stamatakis *et al.*, 2005) algorithm implemented in ARB using default settings.

The metagenome sequences and the mRNA fraction of the metatranscriptome were functionally annotated for initial screening using the metagenomics (MG)-RAST server with maximum *e*-value for a significant match set to 1e − 4 (Meyer *et al.*, 2008). The sequences were also taxonomically binned by MEGAN analysis (parameters: min. bit score 50, min. support 1, top percent 2; 50 best blast hits) of BLASTX files against the RefSeq protein database (*e*-value <1e − 4). Screening for specific functional genes and transcripts were carried out using custom reference databases generated from the UniProtKB/Swiss-Prot or UniProtKB/TrEMBL databases. BLASTX searches were carried out with an *e*-value threshold of 1e − 10. All sequences below threshold were selected as query sequences for a BLASTX search with the same parameters against

the RefSeq database. Only those that were assigned the same function in the second round were considered positives.

In the metagenomes, the genes encoding carbohydrate-active enzymes were annotated. The metagenomic reads were translated into all six frames, each frame into separate ORFs (open reading frames), avoiding any '*' characters marking stop codons in a resulting ORF. All ORFs equal to 40 amino acids or larger were screened for assignable conserved protein domains. Glycoside hydrolases (GH), proteins with a carbohydrate-binding motif, carbohydrate esterases, polysaccharide lyases and other carbohydrate-active enzymes were inspected by reference HMMs (Hidden Markov Models) using HMMER tools (http://hmmer.janelia.org/) with the PFAM database HMMs (PFAM release 25, http://pfam.janelia.org). All database hits with *e*-values below a threshold of 10–4 were counted. For the GH44 family, no PFAM HMM is available. Therefore, representative sequences were selected from the CAZy website (http://www.cazy.org), the sequence regions corresponding to the family were determined and used to create a HMM for the GH44 family. The screening was performed on the HPC computer STALLO at the University of Tromsø (http://docs.notur.no/uit). A selection of sequences was also used as query sequences in BLASTP searches against the RefSeq database of the NCBI. Within the chosen threshold of the HMMER, all identified carbohydrate-active enzymes gave either a corresponding hit or no hit in the RefSeq database. For the taxonomic assignment of gene and transcript sequences, the corresponding BLAST outputs were uploaded in MEGAN (parameters: min. bit score 35, min. support 1, top percent 10; 50 best blast hits).

Statistical analysis
Significant differences between the frequencies of conserved protein domains in ORFs of different soils were evaluated statistically by using the R package (R_Development_Core_Team, 2009) using the χ^2 contingency table test. The contingency table holds the frequency counts of hits and non-hits for a certain PFAM domain category of two different soils. The total frequency count is given by all hits found for any domain in the PFAM database. In cases where the frequencies are too low to meet the rules of the test, the probabilities of observing a sample statistic as high as the test statistic were calculated by Monte Carlo simulations with 100 000 replicates.

Data deposition
The sequence data generated in this study was deposited in the Sequence Read Archive of NCBI and are accessible through accession number SRP014474.

Results and discussion

The active layers of the two permafrost fens at Solvatn and Knudsenheia were slightly acidic

(pH5–6), had a water content of ∼70–90%, an organic matter content of 40–90% (Supplementary Table S1) and were characterised by a moss cover dominated by *Calliergon richardsonii* (Solheim *et al.*, 1996; Wartiainen *et al.*, 2003; Høj *et al.*, 2005). The mosses were interspersed by grasses (*Dupontia pelligera)*, which were heavily suppressed by grazing Barnacle geese. High concentrations of acetate, lactate and ethanol indicated anaerobic metabolisms through fermentative and acetogenic pathways, even in the top layers (Supplementary Table S1). Remarkably high yields of nucleic acids (DNA and RNA) per gram dry peat were obtained from all samples and layers (Supplementary Table S2), which suggests a high microbial abundance and activity at the peak of the growing season. The yields, especially of RNA, decreased with depth, indicative of decreasing microbial activity. Replicated metatranscriptome data sets were obtained from all upper layers (Supplementary Figure S3) whereas the reverse transcription was severely inhibited in the lower layers, presumably owing to phenolic substances (Supplementary Table S2). Metagenomic libraries were obtained from all layers of one duplicate from Solvatn and Knudsenheia (K1 and S2, Supplementary Figure S3). 454 Titanium pyrosequencing of the metatranscriptomes and metagenomes resulted in 120 000–190 000 sequences per library, with one exception (S2b complementary DNA, Supplementary Table S3). The number of rRNA sequences in these libraries ranged from 100 000 to 150 000 while the number of putative mRNA sequences ranged from 7000 to 8500.

Community structure of active microorganisms
The community composition of bacteria, archaea and eukaryotes, as determined from the small subunit rRNA fragments (SSU ribo-tags) of the metatranscriptomes (up to 74 000 ribo-tags), revealed that the biological replicates of both sites had a similar composition of taxa at domain, phylum and class levels of resolution (Supplementary Figure S4). Ribo-tags are considered a measure of living biomass, as they stem from ribosomes, most likely indicative of active organisms (Urich *et al.*, 2008; Urich and Schleper, 2011). Bacteria comprised most ribo-tags; ∼70%–80% in all upper layers and ∼84% in the lower, anoxic layer of Solvatn, S2b (Figure 1). Eukaryotic ribo-tags comprised ∼15%–30%, whereas the fraction of Archaea was surprisingly small in all the upper layers (∼0.01%–0.13%), but higher in S2b (∼1%). The majority of archaeal sequences stemmed from methanogenic archaea. Similar low relative abundances of Archaea have been reported from Alaskan permafrost soils (for example, Waldrop *et al.*, 2010; Wilhelm *et al.,* 2011). The bacterial community composition in both peat soils was very similar, with most ribo-tags stemming from eight phyla, most

notably the Proteobacteria (∼37%–45% of bacterial ribo-tags), Actinobacteria (∼15%), Planctomycetes (∼9%–14%), Verrucomicrobia (∼9%–11%), Acidobacteria (∼8%–10%) and the Chloroflexi (∼7%–9%). Within the Proteobacteria, Deltaproteobacteria (mainly *Myxococcales*) were the most abundant class (∼19%–21% of bacterial ribo-tags). Alphaproteobacteria made up ∼8%, Betaproteobacteria comprised ∼5% while Gammaproteobacteria represented on average 3.2% of the total community. The major eukaryotic taxon was the Protist kingdom Alveolata, putative grazers of bacteria, which represented ∼17%–20% of eukaryotic ribo-tags in both layers of Solvatn and 6%–8% in the top layer of Knudsenheia (Figure 1). Other abundant Protists were the Amoebozoa, Rhizaria (both predators of bacteria) and Stramenopiles. Interestingly, the Protists constituted the major eukaryotic group, exceeding even the fraction of fungal (∼2%–8%) and metazoan (∼11%–13%) ribo-tags (Figure 1). This is in contrast to temperate and boreal mineral soils, where the fungi and metazoa comprise the majority of either eukaryotic ribo-tags (Urich *et al.*, 2008) or biomass (Schaefer, 1990; Berg and Bengtsson, 2007) and the Protists are much less abundant. Our data indicate that the bacterial energy channel might be more important than the fungal energy channel in these peat soils. Metazoan ribo-tags originated from typical micro- and meso-faunal groups of higher trophic levels, including Nematoda, Tardigrada, Rotifera and Arthropoda. Ascomycota, and to a minor extent Basidiomycota, dominated the small fungal community. At Solvatn, a change in abundance with depth was observed for several taxa (Figure 1). While Alphaproteobacteria, Betaproteobacteria and Chloroflexi were present in larger fractions in the lower layer, Planctomycetes and Verrucomicrobia decreased with depth. Most notably, however, was the increase in the fraction of Archaea and the decrease of Protists and fungi with depth. The ribo-tags originated mostly from actively transcribing organisms, indicated by a taxonomic binning of the mRNA fraction, which showed similar community profiles for the bacterial taxa (Figure 2a).

There are few studies addressing Arctic peat soils to compare our metatranscriptomic data with. 16S rRNA gene clone library, and metagenomic studies of permafrost mineral soils have shown that the Actinobacteria, Proteobacteria and Acidobacteria dominate these environments (Hansen *et al.*, 2007; Liebner *et al.*, 2008; Steven *et al.*, 2008; Chu *et al.*, 2010; Yergeau *et al.*, 2010; Wilhelm *et al.*, 2011), while clone library, T-RFLP and FISH studies of a sub-Arctic acidic sphagnum peat report a dominance of Alphaproteobacteria, Betaproteobacteria, Acidobacteria, Verrucomicrobia and Planctomycetes (Dedysh *et al.*, 2006; Pankratov *et al.*, 2011). A recent study of permafrost peat showed that the active layer communities were dominated by Actinobacteria, Proteobacteria and Chloroflexi, while the permafrost microbiota also contained large
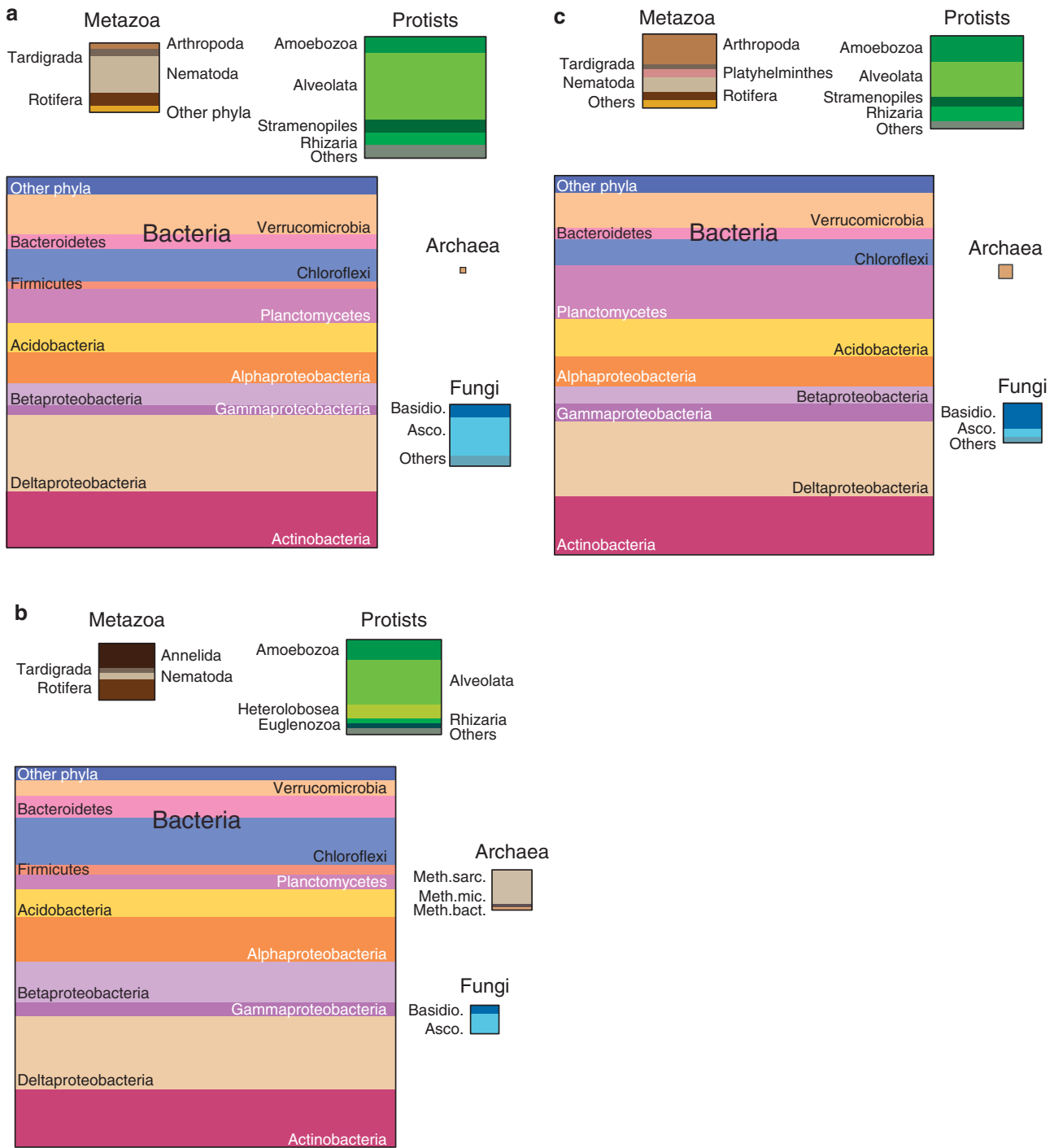
Figure 1 Three-domain community profile of the microbiota in Svalbard peatlands. The figures are based on the ribo-tags fraction of the metatranscriptome. The size of the boxes is proportional to the fraction of ribo-tags of the respective taxa. (**a**) Top layer of Solvatn peat generated from two biological replicate data sets (S1a and S2a). (**b**) Lower layer of Solvatn generated from one data set (S2b). (**c**) Top layer of Knudsenheia peat generated from two biological replicate data sets (K1a and K2a). Asco, Ascomycota; Basidio, Basidiomycota; Meth.bac., *Methanobacteriales*; Meth.mic., *Methanomicrobiales;* Meth.sarc., *Methanosarcinales*.

populations of Bacteroidetes and Firmicutes (Mackelprang *et al.,* 2011). Interestingly, taxa such as the Verrucomicrobia and Planctomycetes, which were among the most abundant taxa in the Svalbard peat soils as well as the acidic Sphagnum peat, were not reported as major taxa in the studies on mineral soils.

Genomic potential for degradation of plant polymers
Plant polymer and phenolic compound degradation are key processes of SOC decomposition in peat soils. We analysed the genomic potential for these processes, by screening the metagenomes for genes encoding conserved protein family (PFAM) domains
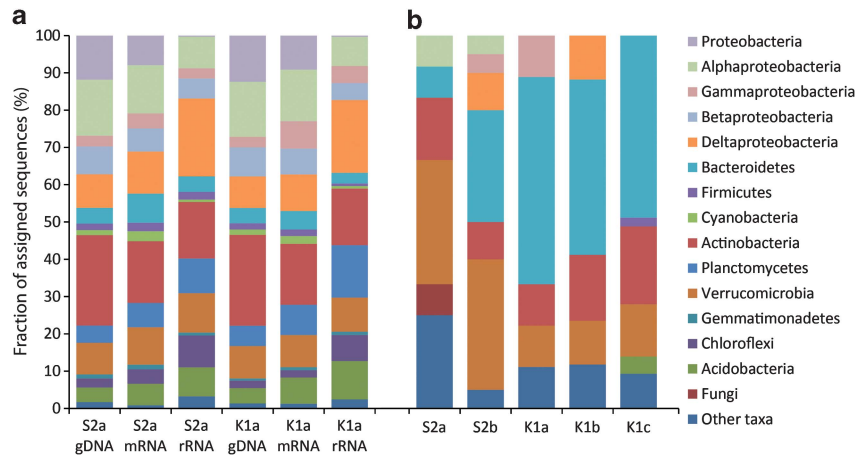
Figure 2 Taxonomic assignment of metatranscriptomic and metagenomic sequences. The community structure is displayed at phylum level resolution with Proteobacteria split into classes. Sequences assigned to Proteobacteria refer to sequences that could not be assigned to class level resolution. S2 and K1 indicates where the samples were collected, in Solvatn and Knudsenheia respectively, while a, b and c indicates the depth of the sample from top and down (Supplementary Figure S3). (**a**) Taxonomic assignment of ribo-tags, mRNA and metagenomic DNA (gDNA) to the domain bacteria. (**b**) Taxonomic assignment of the metagenomic sequences encoding polysaccharide-degrading enzymes (cellulases, endohemicellulases and debranching enzymes). All sequences assigned to these categories were pooled together and taxonomically binned using MEGAN (see Materials and methods for details).

of relevant enzymes, that catalyse the hydrolysis of abundant plant polymers such as cellulose and hemicelluloses, and the degradation of phenolic compounds (for example, lignin) (Table 1). Despite the rather uniform plant cover, dominated by the moss *C. richardsonii,* we identified a large variety of CAZy (carbohydrate-active enzyme) families, including 76 GH families, 36 carbohydrate-binding module, 3 phenol oxidase, 3 polysaccharide lyase and 5 carbohydrate esterase families (Supplementary Table S4). GH5 were the most abundant among the families of cellulases. Genes encoding endohemicellulases (especially GH families 10 and 26 that target xylan), debranching enzymes and phenol oxidases involved in the degradation of phenolic compounds were also abundant (Table 1).

We wanted to know if the microbiotas of our moss-dominated, carbon-accumulating high-Arctic peat soils contain a different genetic potential for the degradation of plant polymers and recalcitrant phenolic compounds than the microbiota of soils from other climate zones, having a vascular plant cover. Therefore, we compared our metagenomic data sets with metagenomes from temperate grasslands and farmlands, subtropic rain forest soils and Arctic permafrost mineral soils (see Supplementary Table S5 for details about the soil data sets), focussing on the PFAMs relevant for plant polymers and phenolic compound degradation. The functional potential was similar in all soils, and genes for the same PFAMs dominated the metagenomes (Supplementary Table S6). A χ^2 test did not detect statistically significant differences between the metagenomes from arctic peat and the other metagenomes (Supplementary Table S7), regardless of the variations in geography or plant coverage.

The chemical composition of moss cell walls is generally considered to be simpler than the one from vacular plants; mosses have a different and less complex hemicellulose composition, and do not contain lignin, but a simpler and less abundant phenolic compound called lignan (Pena *et al.*, 2008; Sarkar *et al.*, 2009; Popper *et al.*, 2011), with the exception of *Sphagnum* species containing an abundant lignin-like polymer (Ligrone *et al.*, 2008). It was therefore surprising that the microbiotas harboured such a similar genomic potential. This result might to some extent be explained by methodological constraints. Fon instance, the PFAMs for phenolic compound degradation do not distinguish between lignin and lignan as substrate. Also, the enzymes themselves might have a broad substrate specificity. Finally, the presence of many similar types of polymers in both mosses and vascular plants should be reflected in an overall similar genetic repertoire for degradation in the respective soil microbiotas. It might well be that the actual composition of plant polymers is only reflected in the gene expression patterns of the microbiota and not in its genetic potential. However, the rather small number of mRNAs in the metatranscriptome prevented us from studying this.

Another major constraint for SOC degradation, besides the short growing season, is the accumulation of phenolic compounds in the water-logged, predominantly anaerobic peat soils, which inhibits enzymatic activity (Fenner and Freeman, 2011). Expression of phenol oxidases was detected in the top layer of Solvatn. However, the genetic potential for phenol oxidases was observed in all peat layers, indicating that a degradation of these inhibitory compounds can occur if the peat soils get oxygenated, possibly through more pronounced drought-rewetting cycles at higher temperatures, like it has recently been shown for temperate peatlands (Fenner and Freeman, 2011).

Table 1 PFAMs that target plant polymers, identified in metagenomes from high-Arctic peatlands of Svalbard

	S2a	S2b	K1a	K1b	K1c
Cellulases					
GH5	4 (2.3)	4 (1.8)	4 (2.1)	8 (3.9)	18 (5.8)
GH6	2 (1.2)	0 (0)	1 (0.5)	1 (0.5)	2 (0.7)
GH7	0 (0)	1 (0.5)	0 (0)	1 (0.5)	0 (0)
GH9	1 (0.5)	2 (0.9)	4 (2.1)	1 (0.5)	6 (2.0)
GH44	2 (1.2)	2 (0.9)	3 (1.6)	2 (1.0)	0 (0)
GH45	1 (0.6)	0 (0)	0 (0)	0 (0)	1 (0.3)
GH48	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Sum	10 (5.9)	9 (4.1)	12 (6.3)	13 (6.3)	27 (8.8)
Debranching enzymes					
GH51	5 (2.9)	8 (3.6)	3 (1.6)	5 (2.4)	13 (4.2)
GH54	1 (0.6)	3 (1.4)	0 (0)	0 (0)	0 (0)
GH62	0 (0)	0 (0)	1 (0.5)	0 (0)	0 (0)
GH67	2 (1.2)	1 (0.5)	0 (0)	0 (0)	2 (0.7)
GH78	1 (0.6)	2 (0.9)	2 (1.1)	10 (4.9)	15 (4.9)
Sum	9 (5.3)	14 (6.4)	6 (3.1)	15 (7.3)	30 (9.7)
Endohemicellulases					
GH8	0 (0)	0 (0.0)	1 (0.5)	0 (0)	2 (0.7)
GH10	2 (1.2)	6 (2.7)	1 (0.5)	5 (2.4)	7 (2.3)
GH11	0 (0.0)	0 (0.0)	0 (0.0)	0 (0)	1 (0.3)
GH12	0 (0.0)	1 (0.5)	0 (0.0)	0 (0)	0 (0)
GH26	4 (2.3)	4 (1.8)	0 (0)	1 (0.5)	12 (3.9)
GH28	0 (0.0)	1 (0.5)	4 (2.1)	2 (1.0)	8 (2.6)
GH53	0 (0.0)	3 (1.4)	3 (1.6)	2 (1.0)	4 (1.3)
Sum	6 (3.5)	15 (6.8)	9 (4.7)	10 (4.9)	34 (11.0)
Oligosaccharide-degrading enzymes					
GH1	9 (5.3)	10 (4.6)	6 (3.1)	14 (6.8)	20 (6.5)
GH2	5 (2.9)	0 (0.0)	4 (2.1)	4 (1.9)	15 (4.9)
GH3	5 (2.9)	9 (4.1)	6 (3.1)	9 (4.4)	18 (5.8)
GH29	5 (2.9)	4 (1.8)	4 (2.1)	11 (5.3)	13 (4.2)
GH35	2 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.0)
GH38	0 (0.0)	0 (0.0)	1 (0.5)	3 (1.5)	4 (1.3)
GH39	2 (1.2)	5 (2.3)	1 (0.5)	2 (1.0)	7 (2.3)
GH42	3 (1.8)	2 (0.9)	2 (1.1)	6 (2.9)	11 (3.6)
GH43	8 (4.7)	7 (3.2)	5 (2.6)	8 (3.9)	10 (3.2)
GH52	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.0)	0 (0.0)
Sum	39 (23)	37 (17)	29 (15)	59 (29)	101 (33)
Phenolic compound-degrading enzymes					
Laccase	1 (0.6)	8 (3.6)	7 (3.7)	11 (5.3)	12 (3.9)
Dioxygenase	2 (1.2)	8 (3.6)	3 (1.6)	2 (1.0)	5 (1.6)
Peroxidase	4 (2.3)	5 (2.3)	3 (1.6)	8 (3.9)	1 (0.3)
Sum	7 (4.1)	21 (9.6)	13 (6.8)	21 (10.2)	18 (5.8)
Tot. ass. seq.	17 067	21 980	19 104	20 581	30 834

The metagenomic profile is divided into five categories according to the functional role of the protein families in plant polymer degradation, a format previously applied in Allgaier *et al.* (2010) and Pope *et al.* (2010). Numbers reflect the number of significant hits (see Materials and methods for details). Shown in brackets are the numbers in $^o/_{ooo}$ of the total number of sequences assigned to a PFAM. Tot. ass. seq., the number of sequences that were assigned to a PFAM (See Materials and methods for details). See Supplementary Table S4 for PFAM accession numbers and description of protein families.

We aimed to identify the microorganisms possessing the genes encoding these hydrolytic enzymes by using taxonomic binning with the MEGAN software (See Materials and methods for details). However, this type of taxonomic binning is problematic, owing to the unequal representation of reference genomes for many taxa in the public databases (Urich *et al.*, 2008) and horizontal gene transfer. We therefore compared the community profiles from SSU ribo-tags with the taxonomically assigned metagenomic and metatranscriptomic sequences from the peat top layers using MEGAN (Figure 2a).

These profiles were similar for many taxa, suggesting that genes and transcripts can be used as taxonomic markers with some level of confidence. The genes assigned to the three major categories of polysaccharide degradation (cellulases, endohemicellulases and debranching enzymes; see Table 1) were taxonomically annotated. The bacterial phyla Bacteroidetes, Actinobacteria and Verrucomicrobia possessed the majority (>70%) of these genes (Figure 2b). The profiles of the two peat soils differed, with the Knudsenheia metagenomes having a higher fraction of genes from Bacteroidetes and

306

Table 2 Metagenomic and metatranscriptomic reads assigned to enzyme classes that catalyse key steps in anaerobic respiration and fermentation, and fermentative pathways

Pathway/function	E.C.	Gene	Metagenomes					Metatranscriptomes	
			S2a	S2b	K1a	K1b	K1c	S1a + S2a	K1a + K2a
Nitrate reduction	1.9.6.1	*napA*	1	6	7	4	1	0	0
Nitrate reduction	1.7.99.4	*narG*	10	6	14	19	28	0	0
Nitrite reduction	1.7.2.1	*nirK*	1	1	2	3	3	0	0
Nitrite reduction	1.7.2.1	*nirS*	0	1	0	2	1	0	2
Nitric oxide reduction	1.7.99.7	*norB*	1	4	3	2	4	0	1
Nitrous oxide reduction	1.7.99.6	*nosZ*	3	3	1	2	1	0	0
Sulphate reduction	1.8.99.3	*dsrA*	0	1	1	4	11	0	0
Fermentation: [FeFe] hydrogenase	1.12.7.2	*hydA*	2	9	0	7	8	0	0
Homoacetogenesis	6.3.4.3	fthfs	6	12	7	10	23	0	0
Methanogenesis	2.8.4.1	*mcrA*	0	0	1	0	2	0	0
Acetoin, butanediol metabolism			55	92	60	85	133	8	8
Acetone, butanol, ethanol synthesis			243	298	319	280	346	17	17
Acetyl-CoA fermentation to butyrate			310	362	406	329	407	16	22
Butanol biosynthesis			235	293	315	263	308	9	19
Fermentations: lactate			52	76	56	74	104	2	3
Fermentations: mixed acid			107	157	121	131	190	5	8
Tot. ass. sequences			69 176	63 527	85 664	54 889	72 166	6315	6205

First column shows function or main pathway, second column shows the enzyme class entry of the ExPASY bioinformatics resource, while third column shows the gene abbreviation. Merged biological replicates for the two sites Solvatn (S1a + S2a) and Knudsenheia (K1a + K2a) are listed under metatranscriptomes. Total ass. sequences, the number of sequences assigned a function by MEGAN analysis. Sequences assigned to pathways were annotated using (MG)-RAST (see Materials and methods for details).

Actinobacteria, while Solvatn had a higher fraction of genes assigned to Verrucomicrobia. Surprisingly, few genes were taxonomically assigned to fungi (Figure 2b). This was in accordance with the low abundance of fungal, and especially basidiomycotal ribo-tags (Figure 1). The latter have been reported as particularly potent decomposers of SOC (Waldrop *et al.*, 2010). Low fungal abundances in permafrost soils and the influence of plant community composition on fungal population size have been shown before (Zak and Kling, 2006; Yergeau *et al.*, 2010; Waldrop *et al.*, 2010).

Anaerobic respiration and fermentation
We investigated the genomic potential for and the expression of key enzymes in anaerobic respiration and fermentation, processes which drive the anaerobic decomposition of SOC. Especially the genes encoding dissimilatory nitrate reductases were abundant (Table 2). This suggests that denitrification is important, which is supported by the high nitrate concentrations detected (70 and 84 µg per gram soil in Solvatn and Knudsenheia, respectively; Alves, 2011). An important role of denitrification has also been suggested by recent metagenomic studies of permafrost affected soils (Yergeau *et al.*, 2010; Mackelprang *et al.*, 2011). Despite the high abundance of denitrification genes, only *nirS* (nitrite reductase) and *norB* (nitric oxide reductase) transcripts were identified (Table 2); however, the small size of the mRNA pools in the metatranscriptomic data sets prevented further statements. Also sulphate reduction might have a role, as genes of the key enzyme DSR were found and their abundance

increased with soil depth in Knudsenheia (Table 2). The majority of the identified anaerobic respiration genes were assigned to taxa within Actinobacteria and Proteobacteria (Supplementary Figure S5), although the genes were broadly distributed among several different taxa, as previously described for denitrifyers (Philippot and Hallin, 2005) and sulphate reducers (Wagner *et al.*, 2005). Genes and transcripts of several fermentative pathways were detected (Table 2), explaining the high concentration of fermentation products in the peat soils (Supplementary Table S1). Likewise, genes encoding the terminal hydrogenase of H_2-evolving fermentations (*hyd*A) were found. Their abundance increased with depth, as did the abundance of formyltetrahydrofolate synthetase genes (*fhs*), which encodes the key enzyme of the Acetyl-CoA pathway of homoacetogenesis. The majority of *hyd*A and *fhs* genes were assigned to Firmicutes and Actinobacteria, suggesting an important role of taxa within these phyla in the anaerobic degradation of SOC in the Svalbard peat (Supplementary Figure S6). Taken together, these findings indicate a gradual transition from aerobic to anaerobic degradation of SOC with depth.

Methane cycling
The fraction of methanogenic archaea increased with depth (Figure 1 and Supplementary Figure S6), reflecting the presumably more anoxic conditions and higher concentrations of fermentation intermediates (Supplementary Table S1). Three orders of methanogenic archaea were found, *Methanobacteriales*, *Methanomicrobiales* and *Methanosarcinales*,
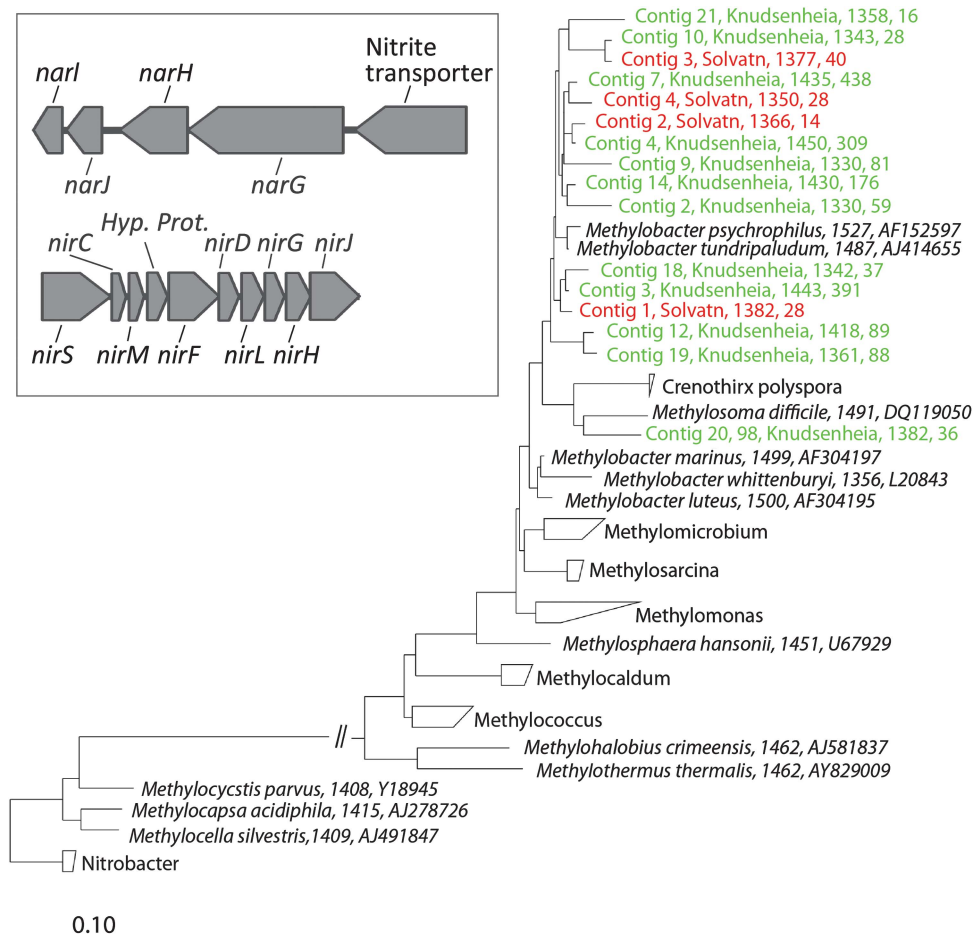
Figure 3 *M. tundripaludum* in Svalbard peat soils. The phylogenetic tree shows assembled SSU rRNA contigs of type I methanotrophs. Most of the nearly full-length ribo-contigs (14 out of 16) are >97% identical to *M. tundripaludum*. Ribo-contig description includes the following: contig id, site (Knudsenheia-green, Solvatn-red), contig length and the number of single ribo-tags that went into the assembly. The reference sequence description (black) includes sequence length and accession number. The length of the bar indicates 0.10 changes per nucleotide. The tree was constructed using the ARB software (See Materials and methods for details). The insert shows the nitrate reductase (open reading frames (ORFs) 3936–3940) and nitrite reductase operons (ORFs 3936–3940) identified in the genome of *M. tundripaludum*.

Table 3 mRNAs assigned to *M. tundripaludum* in the Knudsenheia metatranscriptome

Annotation	AC	Function	No. of reads
pmoC	ZP_07656564.1	Methane oxidation	8
pmoA	ZP_07656563.1	Methane oxidation	7
pmoB	ZP_07656562.1	Methane oxidation	6
CsbD family protein	ZP_07656622.1	Stress response	2
MgtE intracellular region	ZP_07654467.1	Magnesium transport	2
Transport-associated protein	ZP_07656600.1	Transport	2
Methanol dehydrogenase	ZP_07652313.1	Methanol oxidation	1
6-Phospho 3-hex.	ZP_07653633.1	RuMP pathway step 2	1
Succinate dehydrogenase	ZP_07652392.1	Electron transport Complex II	1
NADH dehydrogenase	ZP_07653777.1	Electron transport Complex I	1
NADH:ubiquinone oxidoreductase	ZP_07655325.1	Electron transport Complex I	1
Elec. trans. prot. SenC	ZP_07652162.1	Biogenesis of respiratory systems	1
Nitrogenase iron protein	ZP_07654124.1	Nitrogen fixation	1

Abbreviations: elec. trans. prot. SenC, electron transport protein SenC; mRNA, messenger RNA; RuMP, ribulose monophosphate.
Functionally annotated transcripts related to energy, carbon and nitrogen metabolism are shown. mRNA sequences were taxonomically assigned to *M. tundripaludum* by MEGAN (see Materials and methods for details).
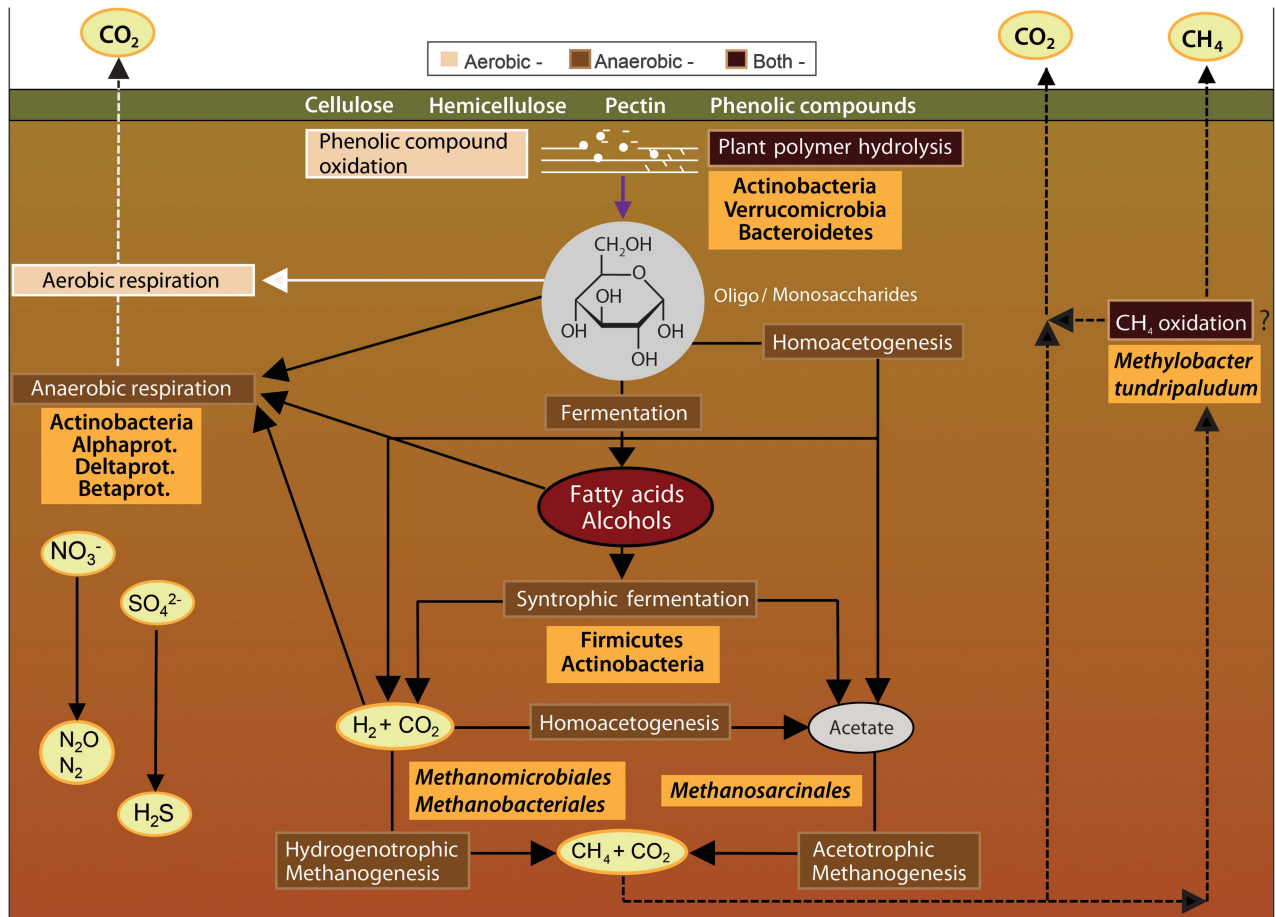
Figure 4 Schematic overview of the main degradation pathways of plant polymers in the high-Arctic peatlands of Svalbard. The pathways are divided into three categories; aerobic (beige), anaerobic (light brown) and processes occurring under both conditions (dark brown). Key microbial taxa for the different degradation steps are presented (orange boxes). The figure is adapted from Figure 1 in (Conrad, 1999).

of which the first two are hydrogenotrophic, producing CH_4 from H_2 and CO_2, whereas the *Methanosarcinales* are metabolically more versatile, carrying out hydrogenotrophic, acetoclastic and methylotrophic methanogenesis. These findings are similar to previous studies of methanogens from Solvatn (Høj *et al.*, 2005; Høj *et al.*, 2006). In Solvatn, the increased abundance of acetotrophic methanogens with depth correlated with a decrease in the concentration of acetate, indicating that the acetotrophic *Methanosarcinales* acted as the sink for acetate. In general, the high concentrations of fermentation products like ethanol, especially in the lower layers of Knudsenheia, indicated a low efficiency in the terminal fermentation steps, possibly owing to low activity of, or a lack of, established interactions between fermentative syntrophic bacteria and their methanogenic counterparts.

Type I methanotrophs of the order *Methylococcales* were abundant in both peat soils (0.1%–2.4% of ribo-tags). We did not detect active type II methanotrophs, anaerobic methanotrophic archaea (ANME), the recently described anaerobic methanotroph *Candidatus Methylomirabilis oxyfera* of the

NC10 candidate phylum (Ettwig *et al.*, 2010) or the Verrucomicrobium *Methylacidiphilium* (Islam *et al.*, 2008). We assembled long, sometimes full-length, SSU rRNAs from the ribo-tags of the type I methanotrophs to determine the phylogenetic position (See Materials and methods (Urich *et al.*, 2008; Radax *et al.*, 2012)). The vast majority of the assembled ribo-contigs was closely related to *M. tundripaludum* (Figure 3), an arctic methanotroph originally isolated from Solvatn (Wartiainen *et al.*, 2006). We could taxonomically assign 104 mRNA tags to genes encoded in the genome of *M. tundripaludum* (Svenning *et al.*, 2011). Although this is a small number, it allowed a superficial view into the gene expression and metabolism of *M. tundripaludum in situ*. Remarkably, as much as 33% of the transcripts encoded subunits of the key enzyme particulate methane monooxygenase (Table 3). This highlights the importance of particulate methane monooxygenase for *M. tundripaludum* cells, as this enzyme catalyses the initial step of both the carbon and energy metabolism through the oxidation of methane to methanol. Transcripts of the methanol dehydrogenase that catalyses the

second step of complete oxidation of methane to CO_2 and the ribulose monophosphate pathway for carbon assimilation, as well as transcripts encoding enzymes of the aerobic respiratory chain, were also identified (Table 3). Recently, *M. tundripaludum* was shown to be the dominating methanotroph in a SIP-RNA experiment with Solvatn peat (Graef *et al.*, 2011). It was also shown to be the dominant methanotroph in two soils from the Siberian and the Canadian Arctic (Liebner *et al.*, 2009; Martineau *et al.*, 2010). Although these are few studies, they point to the importance of methanotrophs closely related to *M. tundripaludum* having a key role in the biological CH_4 filter in Arctic soils. The seemingly ecological advantage of *M. tundripaludum* in these often water-logged Arctic soils might be owing to its N_2-fixation capability (Wartiainen *et al.*, 2006; Svenning *et al.*, 2011), as well as to the presence of nitrate reductase (*nar*) and nitrite reductase (*nir*) operons in its genome (Figure 3) (Svenning *et al.*, 2011). It is tempting to speculate that *M. tundripaludum* could employ the NAR and NIR enzymes in denitrifying anaerobic methane oxidation, utilising a pathway similar to that of *Candidatus M. oxyfera* (Ettwig *et al.*, 2010). Alternatively, it could have the ability to grow anaerobically on simple organic compounds using denitrification. Experiments are underway to test these hypotheses.

Conclusion

We have described the metabolic potential and activity of microorganisms involved in SOC degradation in high-Arctic peat soils (Figure 4). This first combined metagenomic and metatranscriptomic study on any soil (to our knowledge), revealed taxa that are important to specific processes such as hydrolysis of plant polymers (for example, Verrucomicrobia and Bacteroidetes), fermentations (for example, Firmicutes), methanogenesis and methanotrophy. Actinobacteria seemed to be particularly important, having a metabolic potential for carrying out several of the key steps in SOC degradation (Figure 4). Based on their relative abundance, fungi appeared to not have an important role. However, as warming will lead to an extended growing season, this might result in a change of the plant covers from mosses to vascular plants and in more oxygenated active layers, where fungi might become more important for SOC decomposition. Dryer and more oxygenated active layers will then likely lead to increased peat decomposition, owing to lowered concentrations of phenolic substances, caused by increased activity of aerobic microorganisms that synthesise phenol oxidases (Fenner and Freeman, 2011). These combined effects might increase the rate of SOC degradation above that of carbon sequestration, and turn these peat soils into CO_2 sources. Increased active layer depth and higher soil temperatures will likely result in bigger habitat size

for anaerobic bacteria and methanogenic archaea, in the cases where the peat soils remain water-saturated. Here, CH_4 production likely will increase, although this might well be balanced by methanotrophic bacteria, of which *M. tundripaludum* seems to be particularly important. However, to which extent such key populations can adapt to the changing conditions remains a crucial question.

Acknowledgements

References

Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S *et al.* (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **5**: 9.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Alves RJE. (2011). *Ammonia-oxidizing archaea from High Arctic soils.* Master thesis, University of Lisbon: Lisbon, Portugal.

Berg MP, Bengtsson J. (2007). Temporal and spatial variability in soil food web structure. *Oikos* **116**: 1789–1804.

Cao M, Marshall S, Gregson K. (1996). Global carbon exchange and methane emissions from natural wetlands: application of a process-based model. *J Geophys Res* **101**: 399–414.

Chou HH, Holmes MH. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics* **17**: 1093–1104.

Chu HY, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. (2010). Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* **12**: 2998–3006.

Conrad R. (1999). Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiol Ecol* **28**: 193–202.

Coolen MJL, van de Giessen J, Zhu EY, Wuchter C. (2011). Bioavailability of soil organic matter and microbial community dynamics upon permafrost thaw. *Environ Microbiol* **13**: 2299–2314.

Dedysh SN, Pankratov TA, Belova SE, Kulichevskaya IS, Liesack W. (2006). Phylogenetic analysis and in situ identification of bacteria community composition in an acidic Sphagnum peat bog. *Appl Environ Microbiol* **72**: 2110–2117.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MMM *et al.* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.

Fenner N, Freeman C. (2011). Drought-induced carbon loss in peatlands. *Nat Geosci* **4**: 895–900.

Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW *et al.* (2007). Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press: UK.

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.

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

Graef C, Hestnes AG, Svenning MM, Frenzel P. (2011). The active methanotrophic community in a wetland from the high Arctic. *Environ Microbiol Rep* **3**: 466–472.

Hansen AA, Herbert RA, Mikkelsen K, Jensen LL, Kristoffersen T, Tiedje JM *et al.* (2007). Viability, diversity and composition of the bacterial community in a high Arctic permafrost soil from Spitsbergen, Northern Norway. *Environ Microbiol* **9**: 2870–2884.

Harden JW, Mark RK, Sundquist ET, Stallard RF. (1992). Dynamics of soil carbon during deglaciation of the laurentide ice sheet. *Science* **258**: 1921–1924.

Huang XQ, Madan A. (1999). CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–877.

Huson DH, Auch AF, Qi J, Schuster SC. (2007). MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.

Høj L, Olsen RA, Torsvik VL. (2005). Archaeal communities in high Arctic wetlands at Spitsbergen, Norway (78 degrees N) as characterized by 16S rRNA gene fingerprinting. *FEMS Microbiol Ecol* **53**: 89–101.

Høj L, Olsen RA, Torsvik VL. (2008). Effects of temperature on the diversity and community structure of known methanogenic groups and other archaea in high Arctic peat. *ISME J* **2**: 37–48.

Høj 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.

IPCC (2007). Climate Change 2007—The Physical Science Basis Contribution of Working Group I to the Fourth Assessment Report of the IPCC (ISBN 978 0521 88009-1 Hardback; 978 0521 70596-7 Paperback).

Islam T, Jensen S, Reigstad LJ, Larsen Ø, Birkeland N-K. (2008). Methane oxidation at 55 °C and pH 2 by a thermoacidophilic bacterium belonging to the Verrucomicrobia phylum. *Proc Natl Acad Sci USA* **105**: 300–304.

Kotsyurbenko OR. (2005). Trophic interactions in the methanogenic microbial community of low-temperature terrestrial ecosystems. *FEMS Microbiol Ecol* **53**: 3–13.

Kotsyurbenko OR, Friedrich MW, Simankova MV, Nozhevnikova AN, Golyshin PN, Timmis KN *et al.* (2007). Shift from acetoclastic to H-2-dependent methanogenes is in a West Siberian peat bog at low pH values and isolation of an acidophilic Methanobactetium strain. *Appl Environ Microbiol* **73**: 2344–2348.

Lanzen A, Jorgensen SL, Bengtsson MM, Jonassen I, Ovreas L, Urich T. (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol* **77**: 577–589.

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, Rublack K, Stuehrmann T, Wagner D. (2009). Diversity of aerobic methanotrophic bacteria in a permafrost active layer soil of the lena delta, Siberia. *Microb Ecol* **57**: 25–35.

Ligrone R, Carafa A, Duckett JG, Renzaglia KS, Ruel K. (2008). Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga Nitella. *Plant Syst Evol* **270**: 257–272.

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ *et al.* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**: 368–U120.

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* **76**: 5773–5784.

Metje M, Frenzel P. (2005). Effect of temperature on anaerobic ethanol oxidation and methanogenesis in acidic peat from a northern wetland. *Appl Environ Microbiol* **71**: 8191–8200.

Metje M, Frenzel P. (2007). Methanogenesis and methanogenic pathways in a peat from subarctic permafrost. *Environ Microbiol* **9**: 954–964.

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**: 8.

Pankratov TA, Ivanova AO, Dedysh SN, Liesack W. (2011). Bacterial populations and environmental factors controlling cellulose degradation in an acidic Sphagnum peat. *Environ Microbiol* **13**: 1800–1814.

Pena MJ, Darvill AG, Eberhard S, York WS, O'Neill MA. (2008). Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* **18**: 891–904.

Philippot L, Hallin S. (2005). Finding the missing link between diversity and activity using denitrifying bacteria as a model functional community. *Curr Opin Microbiol* **8**: 234–239.

Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA *et al.* (2010). Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci USA* **107**: 14793–14798.

Popper ZA, Michel G, Herve C, Domozych DS, Willats WGT, Tuohy MG *et al.* (2011). Evolution and diversity

of plant cell walls: from algae to flowering plants. In: Merchant SS, Briggs WR, Ort D (eds). *Annual Review of Plant Biology* Vol 62. Annual Reviews: Palo Alto, pp 567–588.

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

R_Development_Core_Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical ComputingVienna. *Austria.*

Radax R, Rattei T, Lanzen A, Bayer C, Rapp HT, Urich T *et al.* (2012). Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community. *Environ Microbiol* **14**: 1308–1324.

Sarkar P, Bosneaga E, Auer M. (2009). Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *J Exp Bot* **60**: 3615–3635.

Schaefer M. (1990). The soil fauna of a beech forest on limestone: trophic structure and energy budget. *Oecologia* **82**: 128–136.

Solheim B, Endal A, Vigstad H. (1996). Nitrogen fixation in Arctic vegetation and soils from Svalbard, Norway. *Polar Biol* **16**: 35–40.

Stamatakis A, Ludwig T, Meier H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456–463.

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.

Svenning MM, Hestnes AG, Wartiainen I, Stein LY, Klotz MG, Kalyuzhnaya MG *et al.* (2011). Genome Sequence of the Arctic methanotroph Methylobacter tundripaludum SV96. *J Bacteriol* **193**: 6418–6419.

Tarnocai C, Canadell JG, Schuur EAG, Kuhry P, Mazhitova G, Zimov S. (2009). Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochem Cycles* **23**: GB2023, 11PP.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554–557.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**: 13.

Urich T, Schleper C. (2011). The 'double RNA' approach to simultaneously assess the structure and function of environmental microbial communities by meta-transcriptomics. In: de Bruijn FJ (ed). *Handbook of Molecular Microbial Ecology*. Wiley-Blackwell: Hoboken, NJ, USA.

Wagner M, Loy A, Klein M, Lee N, Ramsing NB, Stahl DA *et al.* (2005). Functional marker genes for identification of sulfate-reducing prokaryotes. *Methods Enzymol* **397**: 469–489.

Waldrop MP, Wickland KP, White R, Berhe AA, Harden JW, Romanovsky VE. (2010). Molecular investigations into a globally important carbon pool: permafrost-protected carbon in Alaskan soils. *Glob Change Biol* **16**: 2543–2554.

Wartiainen I, Hestnes AG, McDonald IR, Svenning MM. (2006). Methylobacter tundripaludum sp nov., a methane-oxidizing bacterium from Arctic wetland soil on the Svalbard islands, Norway (78 degrees N). *Int J Syst Evol Microbiol* **56**: 109–113.

Wartiainen I, Hestnes AG, Svenning MM. (2003). Methanotrophic diversity in high arctic wetlands on the islands of svalbard (Norway)—denaturing gradient gel electrophoresis analysis of soil DNA and enrichment cultures. *Can J Microbiol* **49**: 602–612.

Wilhelm RC, Niederberger TD, Greer C, Whyte LG. (2011). Microbial diversity of active layer and permafrost in an acidic wetland from the Canadian High Arctic. *Can J Microbiol* **57**: 303–315.

Yergeau E, Hogues H, Whyte LG, Greer CW. (2010). The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. *ISME J* **4**: 1206–1214.

Zak DR, Kling GW. (2006). Microbial community composition and function across an arctic tundra landscape. *Ecology* **87**: 1659–1670.

Supplementary information

Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms

Alexander Tveit, Rainer Schwacke, Mette M Svenning and Tim Urich

Figure S1. Sampling scheme Solvatn.

S1 and S2 refers to biological replicate 1 and 2. 1, 2 and 3 in squares of top scheme and brackets below refers to the 3 subsamples of each biological replicate which were pooled with layer separation (see materials and methods).
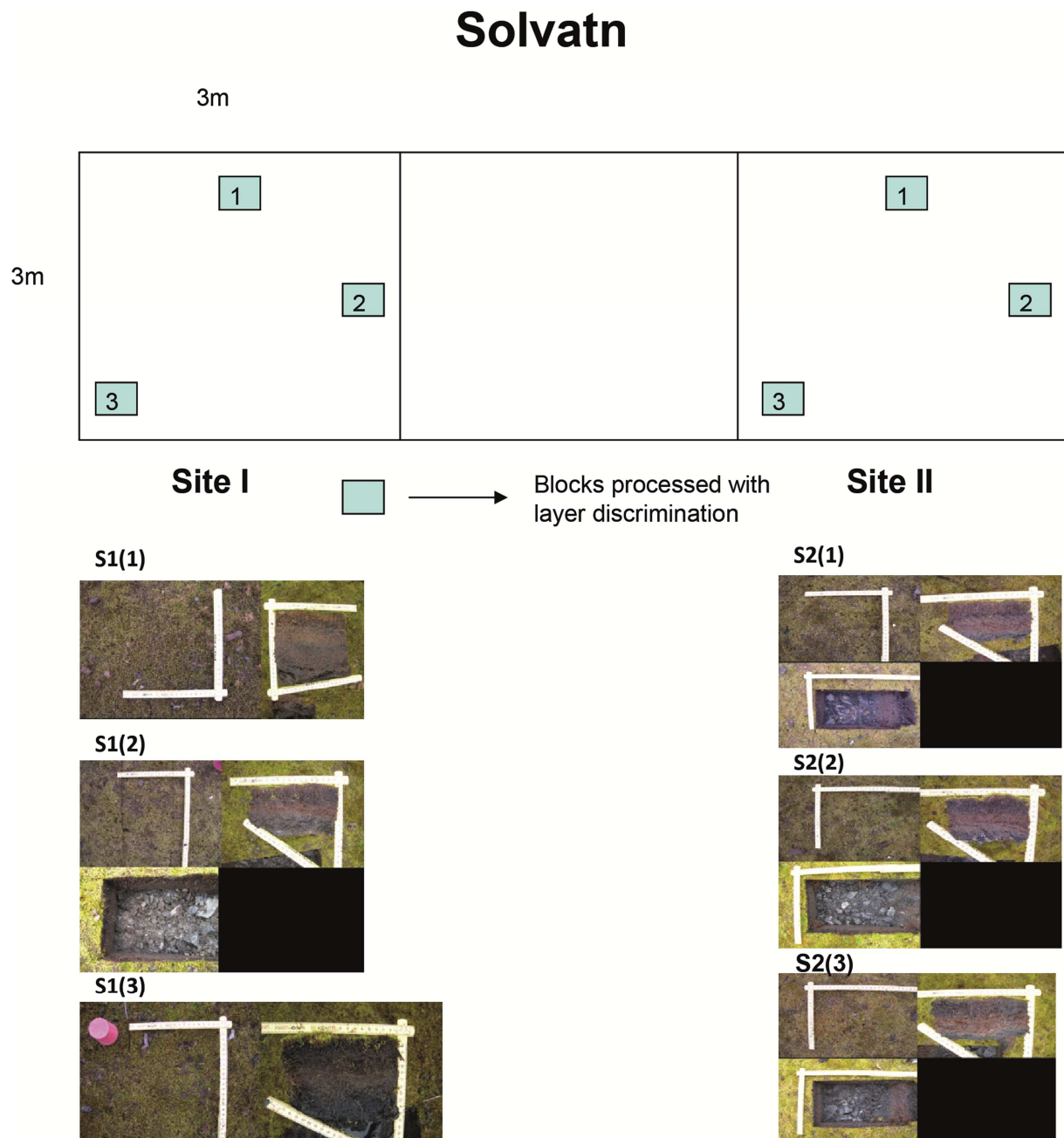
Figure S2. Sampling scheme Knudsenheia.

K1 and K2 refers to biological replicate 1 and 2. 1, 2 and 3 in squares of top scheme and brackets below refers to the 3 subsamples of each biological replicate which were pooled with layer separation (see materials and methods). The third layer that is shown in figure S3, which is not seen in the pictures, refers to the highly decomposed peat that lay beneath the block.
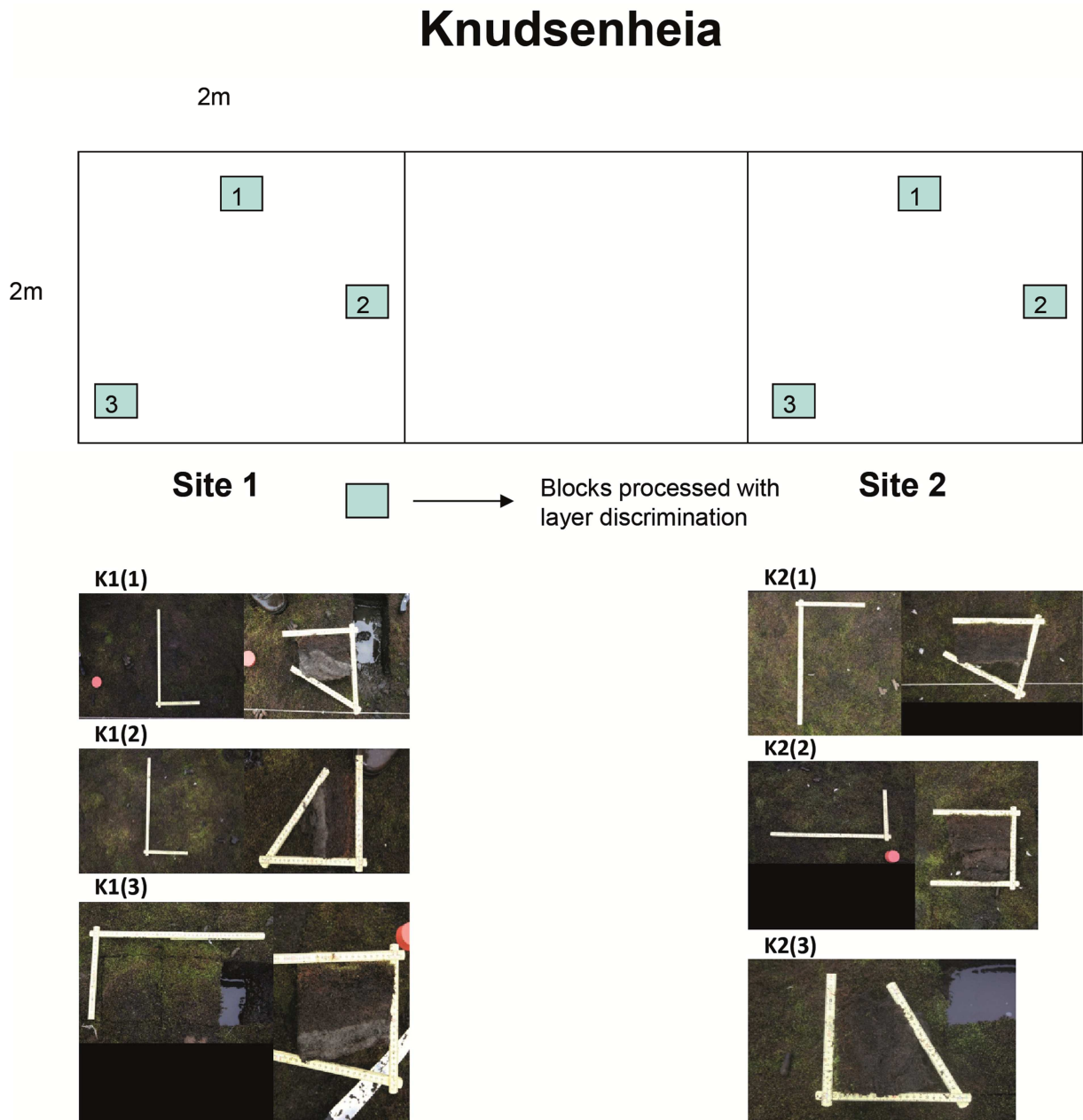
Figure S3. Sampling and analysis scheme

For both sites, Knudsenheia and Solvatn, two biological replicates were prepared, K1 and K2, S1 and S2, each of the biological replicates consisting of three subsamples. a) The two biological replicates from each site are shown, K1 and K2 from Knudsenheia, and S1 and S2 from Solvatn. a, b and c corresponds to the layers of the peat. Datasets were obtained from the layers that are marked: MT: metatranscriptome, MG: metagenome. Only the metatranscriptomes (MT) from the upper layers were analysed with replication. The temperature profiles show the depth of the sampled peat and the corresponding temperatures. b) Sampling and pooling performed to obtain two biological replicates for Knudsenheia (replicates K1 and K2) and Solvatn (replicates S1 and S2). Three peat blocks were sampled from each replicate site (top of figure). The corresponding layers were pooled to generate a merged sample for each layer individually (bottom of figure).
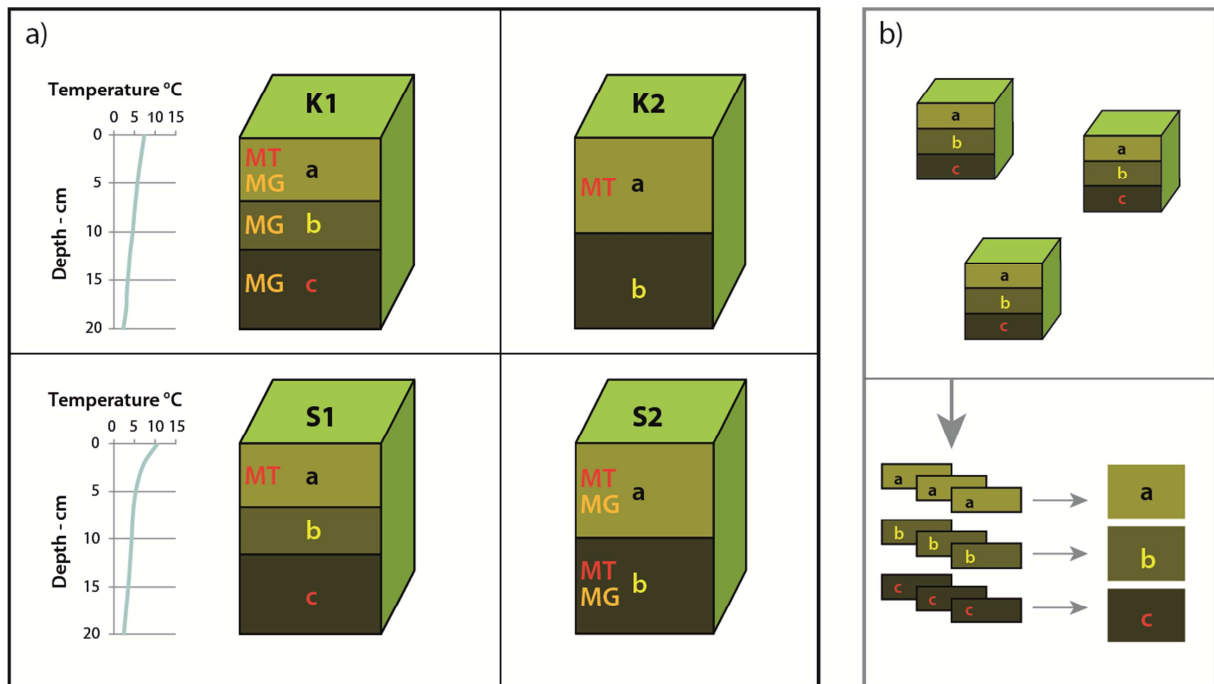
Figure S4. SSU rRNA Community profile comparison between biological replicates of sites Solvatn (S1a, S2a), Knudsenheia (sample K1a, K2a) and the lower layer of Solvatn (S2b).

Column sizes represent the fraction of assigned ribo-tags as a percentage of total assigned reads on domain level (see materials and methods). The bacterial domain is displayed at phylum resolution with proteobacteria split into its respective classes. Taxa making up less than 1% of total number of ribotags were assigned as other phyla/classes.
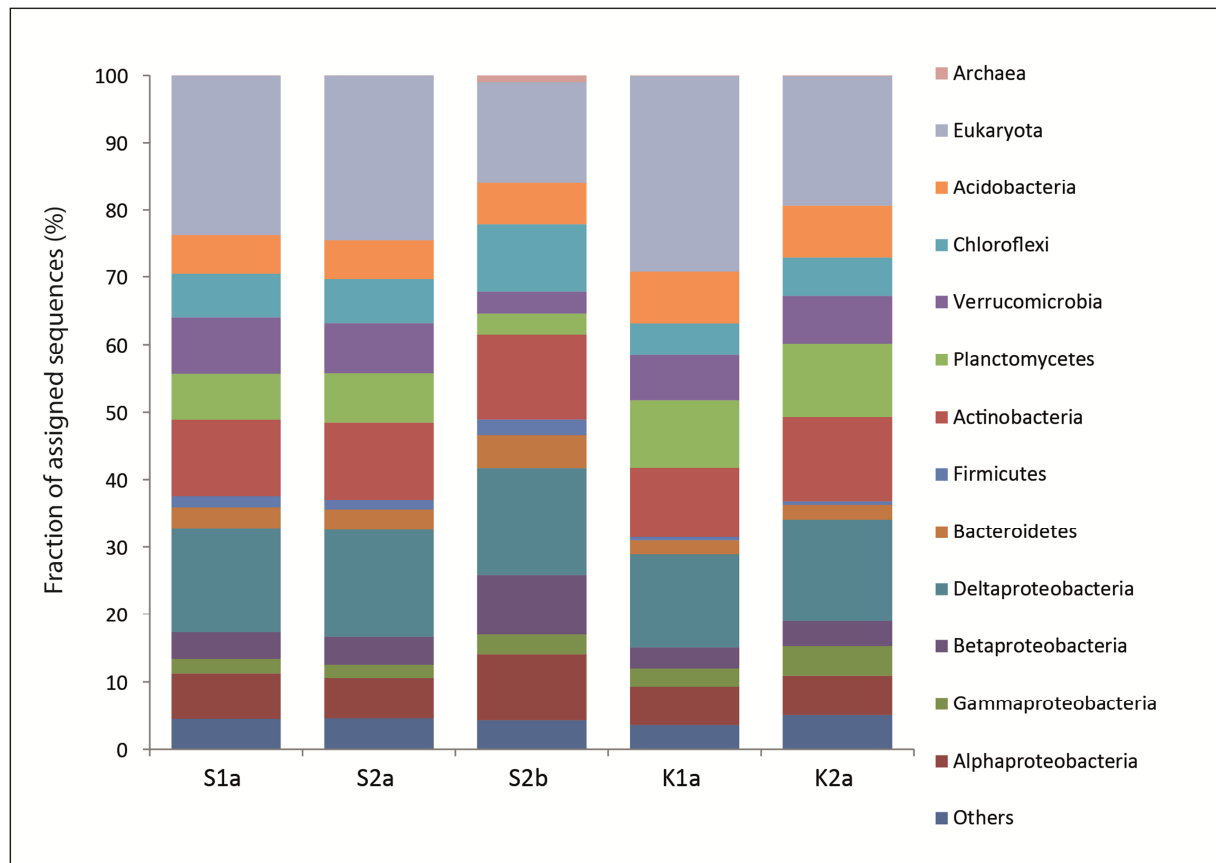
Figure S5. Taxonomic assignment of genes encoding enzymes involved in denitrification (Denitr.), the alpha subunit of the dissimilatory sulfite reductase (*dsrA*), the alpha subunit of the [FeFe] hydrogenase (*hydA*) and the formyl tetrahydrofolate synthetase (fthfs).

Taxonomic binning was carried out by BLASTX searches in the RefSeq database of the NCBI followed by analysis in MEGAN (see materials and methods for details).

The results from the different peat layers were merged due to low number of hits in each layer. Solv.: Solvatn, Knud.: Knudsenheia.
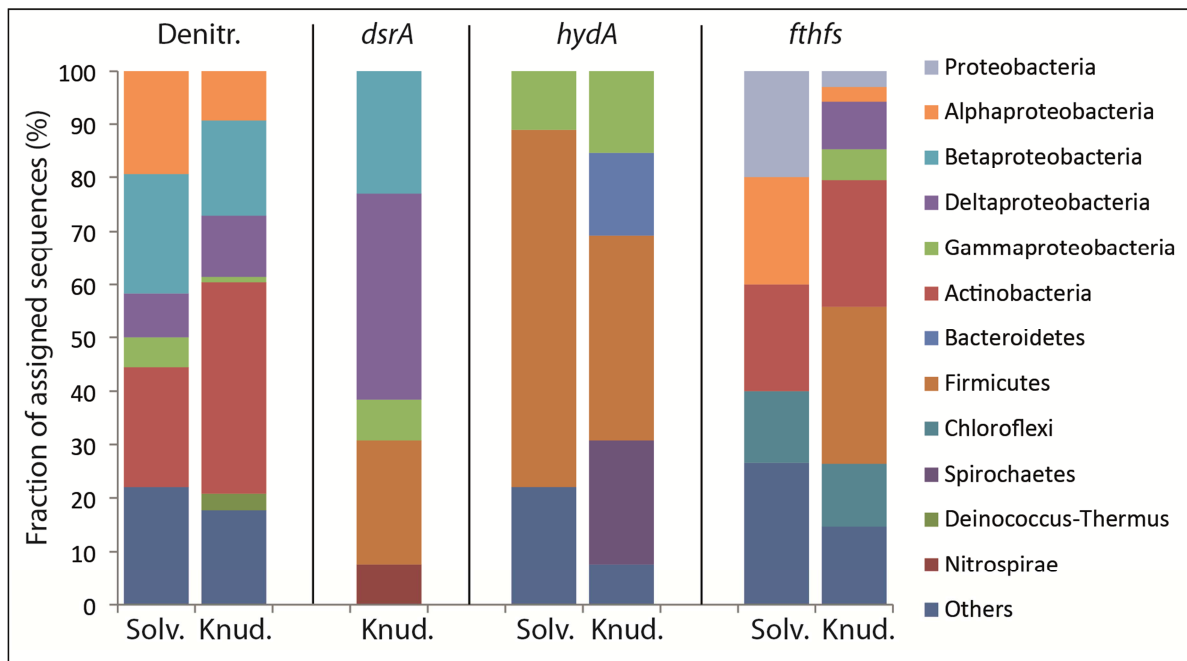
Figure S6. Distribution of methanogens compared between sites and layers.

Results given as a fraction of the total number of sequences assigned to a domain. gDNA: Metagenomic sequences; rRNA: Small subunit rRNA. S2a: Top layer of biological replicate 1, Solvatn; S2b: corresponding lower layer; K1a, b, and c: Top, middle and lower layer of biological replicate 1, Knudsenheia. For K1b and K1c, no metatranscriptomes were generated. Metagenomic sequences were taxonomically assigned at class level resolution, thus no distinction were made between hydrogenotrophic and acetotrophic methanogens. The SSU rRNA was assigned at order level, distinguishing between hydrogenotrophic (Methanobacteriales, Methanomicrobiales) and acetotrophic (Methanosarcinales) methanogens.
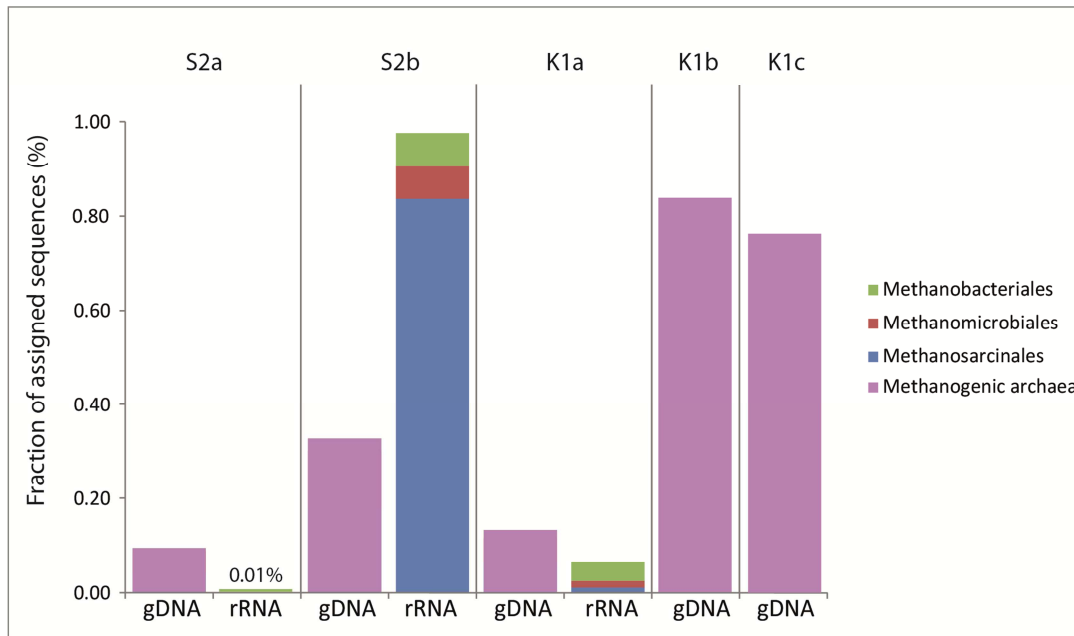
Table S1. Soil characteristics

Sample	pH	Organic matter content (% dry weight)	Water content (% fresh weight)	Lactate (µM)	Acetate (µM)	Ethanol (µM)
S1a	5.4	91	90	194	156	247
S1b	5.3	76	86	16	17	b.d.
S1c	5.0	73	86	12	25	278
S2a	5.2	88	90	91	134	218
S2b	5.2	55	82	13	15	478
K1a	5.8	79	91	50	85	905
K1b	5.4	40	79	11	b.d.[1]	1345
K1c	5.3	55	85	11	43	897
K2a	5.6	76	91	47	49	129
K2b	5.4	71	88	9	41	7731

1: b.d. below detection limit.

7

Table S2. Nucleic acid extraction

Sample	Extracted nucleic acids (μg gdw^{-1})	Spectrophotometric measurement ratio 260/280	Spectrophotometric measurement ratio 260/230	Extracted DNA (μg gdw^{-1})	Extracted RNA (μg gdw^{-1})	RNA/DNA ratio	cDNA synthesis efficiency (%)
S1a	584	2.0	1.8	422	162	0.38	39
S1b	280	1.9	1.5	213	67	0.31	1
S1c	98	1.9	1.0	76	22	0.29	1
S2a	526	1.9	1.9	341	185	0.54	29
S2b	375	2.0	1.8	295	80	0.27	2
K1a	574	2.0	1.7	439	135	0.31	>90
K1b	137	1.9	1.4	111	26	0.23	1
K1c	131	1.9	1.1	111	20	0.18	4
K2a	647	2.0	1.7	512	135	0.26	>90
K2b	125	2.0	1.2	109	16	0.15	1

Table S3: 454 pyro-sequencing statistics

Sample	Total number of pyrosequencing reads	Number of reads after quality and size filtering	Number of ribosomal RNA reads	Number of putative mRNA's	Number of genomic DNA/mRNA reads after removing replicates
S1a cDNA	191164	151477	144095	7382	6642
S2a cDNA	196074	155867	147371	8496	7733
S2a gDNA	120600	97552	-	-	89359
S2b cDNA	4062	2979	2878	101	88
S2b gDNA	148992	118416	-	-	108237
K1a cDNA	185086	145720	137879	7841	7123
K1a gDNA	141451	112917	-	-	103366
K1b gDNA	139464	111727	-	-	101781
K1c gDNA	190057	150467	-	-	136679
K2a cDNA	147018	113569	105158	8411	7640

Table S4. Overview of the protein families searched for in the metagenomes and metatranscriptomes

The numbers displayed are the number of hits in the datasets analysed using the hmmer tool. Accession numbers refer to the corresponding hidden markov models. Protein families that are not represented by a model were not searched for.

Glycoside hydrolase	Accession number		S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rothamsted 0-10	Rothamsted 10-21	Waseca	Eureka active layer	Eureka permafrost
GH1	PF00232.12	Glycosyl hydrolase family 1	9	10	6	14	20	0	0	92	53	43	24	65	11
GH2	PF00703.15	Glycosyl hydrolase family 2	5	0	4	4	15	0	0	24	6	6	11	72	4
GH3	PF00933.15	Glycosyl hydrolase family 3 N terminal domain	5	9	6	9	18	0	0	97	55	39	42	304	28
GH4	PF02056.10	Family 4 glycosyl hydrolase	3	0	1	2	7	0	0	17	12	12	16	18	2
GH5	PF00150.12	Cellulase (glycosyl hydrolase family 5)	4	4	4	8	18	0	0	37	56	22	35	144	13
GH6	PF01341.1	Glycosyl hydrolases family 6	2	0	1	1	2	1	0	2	8	3	5	3	1
GH7	PF00840.14	Glycosyl hydrolase family 7	0	1	0	1	0	0	0	4	4	3	3	7	2
GH8	PF01270.1	Glycosyl hydrolases family 8	0	0	1	0	2	0	0	6	7	5	4	8	1
GH9	PF00759.1	Glycosyl hydrolase family 9	1	2	4	1	6	0	0	31	9	3	13	92	8
GH10	PF00331.14	Glycosyl hydrolase family 10	2	6	1	5	7	0	0	22	28	18	19	68	14
GH11	PF00457.11	Glycosyl hydrolases family 11	0	0	0	0	1	0	0	10	5	5	1	7	1
GH12	PF01670.10	Glycosyl hydrolase family 12	0	1	0	0	0	0	0	2	2	1	4	2	0
GH13	PF00128.18	Alpha amylase, catalytic domain	13	10	10	12	20	0	0	194	135	103	75	634	46
GH14	PF01373.11	Glycosyl hydrolase family 14	0	2	0	0	2	0	0	7	1	2	2	14	0
GH15	PF00723.1	Glycosyl hydrolases family 15	2	5	4	5	7	0	0	76	95	77	23	182	8
GH16	PF00722.15	Glycosyl hydrolases family 16	4	4	3	3	6	0	0	33	25	11	10	123	12
GH17	PF00332.12	Glycosyl hydrolases family 17	0	0	1	3	1	0	1	6	6	2	1	4	1
GH18	PF00704.22	Glycosyl hydrolases family 18	1	1	0	4	4	0	0	19	22	9	18	68	5
GH19	PF00182.13	Chitinase class I	0	0	2	0	0	0	0	2	5	4	5	6	12
GH20	PF00728.16	Glycosyl hydrolase family 20, catalytic domain	2	1	1	4	8	0	0	31	19	12	13	236	26
GH22	PF00062.14	C-type lysozyme/alpha-lactalbumin family	1	0	0	0	0	0	0	2	4	2	2	8	2
GH23	PF01464.14	Transglycosylase SLT domain	10	11	19	7	15	1	0	115	101	62	59	124	18
GH24	PF00959.13	Phage lysozyme	1	0	1	2	0	0	0	3	17	4	1	8	0
GH25	PF01183.14	Glycosyl hydrolases family 25	4	2	2	0	3	0	0	14	12	6	9	16	13
GH26	PF02156.9	Glycosyl hydrolase family 26	4	4	0	1	12	0	1	7	9	9	5	34	2
GH27	PF02065.12	Melibiase	2	3	1	4	7	0	0	26	12	8	11	18	4
GH28	PF00295.11	Glycosyl hydrolases family 28	0	1	4	2	8	0	0	20	10	4	8	62	6
GH29	PF01120.11	Alpha-L-fucosidase	5	4	4	11	13	1	0	68	31	4	23	62	21
GH30	PF02055.10	O-Glycosyl hydrolase family 30	0	0	0	2	5	0	0	15	15	3	10	89	11
GH31	PF01055.20	Glycosyl hydrolases family 31	4	11	8	9	21	0	0	57	23	13	15	164	31
GH32	PF00251.14	Glycosyl hydrolases family 32 N-terminal domain	0	0	1	2	3	0	0	19	14	11	10	132	12
GH33	PF02012.14	BNR/Asp-box repeat	15	18	20	32	40	0	0	355	300	249	157	365	47
GH34	PF00064.12	Neuraminidase	0	0	0	0	1	0	1	6	5	3	2	9	1
GH35	PF01301.13	Glycosyl hydrolases family 35	2	0	0	0	3	0	0	16	16	2	5	4	4
GH37	PF01204.12	Trehalase	1	1	1	1	0	0	0	24	20	4	10	122	14
GH38	PF01074.16	Glycosyl hydrolases family 38 N-terminal domain	0	0	1	3	4	0	0	30	7	7	9	16	3
GH39	PF01229.11	Glycosyl hydrolases family 39	2	5	1	2	7	0	0	21	12	7	15	22	3
GH42	PF02449.9	Beta-galactosidase	3	2	2	6	11	0	0	51	27	15	23	46	8
GH43	PF04616.8	Glycosyl hydrolases family 43	8	7	5	8	10	0	0	62	50	12	46	372	66
GH44	AMB0001.1		2	2	3	2	0	0	0	10	12	3	2	5	0

9

			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenhe ia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
GH45	PF02015.10	Glycosyl hydrolase family 45	1	0	0	0	1	0	1	5	12	4	1	8	1
GH46	PF01374.12	Glycosyl hydrolase family 46	0	0	1	1	0	0	0	3	2	1	1	3	0
GH47	PF01532.1	Glycosyl hydrolase family 47	0	0	0	0	1	0	0	12	10	15	4	13	1
GH48	PF02011.9	Glycosyl hydrolase family 48	0	0	0	0	0	0	0	2	3	1	1	5	0
GH49	PF03718.7	Glycosyl hydrolase family 49	0	0	2	0	0	0	0	3	1	1	1	10	0
GH51	PF06964.6	Alpha-L-arabinofuranosidase C-terminus	5	8	3	5	13	0	0	41	30	8	12	19	6
GH52	PF03512.7	Glycosyl hydrolase family 52	0	0	0	2	0	0	0	3	0	1	1	1	
GH53	PF07745.7	Glycosyl hydrolase family 53	0	3	3	2	4	0	0	13	10	0	6	28	5
GH54	PF09206.5	Alpha-L-arabinofuranosidase B, catalytic	1	3	0	0	0	0	0	5	9	0	2	4	0
GH56	PF01630.12	Hyaluronidase	0	0	1	0	0	0	0	2	5	3	3	8	0
GH57	PF03065.1	Glycosyl hydrolase family 57	3	3	1	5	7	0	0	55	21	14	16	65	4
GH59	PF02057.9	Glycosyl hydrolase family 59	0	1	0	1	0	0	0	7	2	5	4	15	0
GH61	PF03443.8	Glycosyl hydrolase family 61	0	0	0	0	0	0	0	2	2	0	0	2	0
GH62	PF03664.7	Glycosyl hydrolase family 62	0	0	1	0	0	0	0	3	6	3	3	8	0
GH63	PF03200.10	Mannosyl oligosaccharide glucosidase	2	3	0	1	3	0	0	44	50	14	11	35	4
GH65	PF03632.9	Glycosyl hydrolase family 65 central catalytic domain	1	2	5	3	6	0	0	36	27	27	5	120	13
GH67	PF07488.6	Glycosyl hydrolase family 67 middle domain	2	1	0	0	2	0	0	16	3	3	3	18	5
GH68	PF02435.10	Levansucrase/Invertase	0	0	0	0	0	0	0	1	1	2	2	5	0
GH70	PF02324.10	Glycosyl hydrolase family 70	0	2	1	2	0	0	0	4	3	3	6	42	6
GH71	PF03659.8	Glycosyl hydrolase family 71	0	1	1	0	1	0	0	4	3	5	2	5	0
GH72	PF03198.8	Glycolipid anchored surface protein (GAS1)	0	1	0	0	1	0	0	4	6	2	0	10	0
GH73	PF01832.14	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	1	0	3	2	1	0	0	12	10	9	9	69	5
GH75	PF07335.5	Fungal chitosanase	1	1	0	0	0	0	0	5	20	8	0	0	0
GH76	PF03663.8	Glycosyl hydrolase family 76	0	1	1	0	0	1	0	12	5	1	8	33	6
GH77	PF02446.11	4-alpha-glucanotransferase	2	12	7	5	11	0	0	76	86	66	25	242	16
GH78	PF05592.5	Bacterial alpha-L-rhamnosidase	1	2	2	10	15	0	1	59	24	11	15	154	18
GH79	PF03662.8	Glycosyl hydrolase family 79, N-terminal domain	0	1	0	1	0	0	0	3	2	3	3	12	2
GH81	PF03639.7	Glycosyl hydrolase family 81	0	0	0	0	1	0	0	0	2	3	1	6	0
GH83	PF00423.13	Hemagglutinin-neuraminidase	1	0	0	1	0	0	0	4	12	9	2	12	2
GH85	PF03644.7	Glycosyl hydrolase family 85	0	0	0	0	0	0	0	2	2	3	4	8	2
GH88	PF07470.1	Glycosyl Hydrolase Family 88	2	7	5	2	14	0	0	37	34	10	21	79	14
GH89	PF05089.6	Alpha-N-acetylglucosaminidase (NAGLU) tim-barrel domain	0	0	1	0	2	0	0	2	6	1	7	52	4
GH92	PF07971.6	Glycosyl hydrolase family 92	1	4	0	8	9	0	0	39	24	6	7	175	10
GH97	PF10566.3	Glycoside hydrolase 97	4	7	1	3	13	0	0	35	18	3	12	332	29
GH98	PF08306.5	Glycosyl hydrolase family 98	0	0	0	1	2	0	0	0	1	1	0	6	1
GH100	PF04853.6	Plant neutral invertase	0	0	0	1	1	0	0	4	4	4	3	5	2
GH100	PF12899.1	Endo-alpha-N-acetylgalactosaminidase glycoside hydrolase	0	0	0	0	0	0	0	1	1	1	1	4	0
GH102	PF03562.11	MltA specific insert domain	0	1	3	0	0	0	0	12	4	8	4	14	1
GH104	PF00959.13	Phage lysozyme	0	0	0	0	0	0	0	0	0	0	0	0	0
GH105	PF07470.1	Glycosyl Hydrolase Family 88	0	0	0	0	0	0	0	0	0	0	0	0	0
GH108	PF05838.6	Predicted lysozyme (DUF847)	0	3	0	0	1	0	0	8	7	4	2	12	2
Carbohydrate binding module			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenhe ia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
CBM1	PF00734.12	Fungal cellulose binding domain	3	6	4	1	5	1	1	27	39	32	12	19	4
CBM2	PF00553.13	Cellulose binding domain	2	2	6	3	1	0	1	19	31	12	9	15	5

	PF	Description	S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
CBM3	PF00942.12	Cellulose binding domain	2	0	0	0	0	0	0	2	2	3	1	13	4
CBM4	PF02018.11	Carbohydrate binding domain	6	2	1	3	3	0	0	18	15	7	17	26	5
CBM5	PF02839.8	Carbohydrate binding domain	2	1	1	0	1	0	1	12	15	9	5	34	9
CBM6	PF03422.9	Carbohydrate binding module (family 6)	3	1	5	5	9	0	0	24	31	20	10	76	8
CBM9	PF02018.11	Carbohydrate binding domain	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM10	PF02013.10	Cellulose or protein binding domain	0	2	1	1	0	0	0	7	8	3	2	2	1
CBM11	PF03425.7	Carbohydrate binding domain (family 11)	0	1	1	0	1	0	0	6	7	1	6	10	1
CBM12	PF02839.8	Carbohydrate binding domain	0	0	0	0	0	0	1	0	0	0	0	0	0
CBM13	PF00652.16	Ricin-type beta-trefoil lectin domain	6	0	3	1	1	0	0	26	27	7	6	10	0
CBM14	PF01607.18	Chitin binding Peritrophin-A domain	1	3	3	1	3	0	0	11	32	9	13	26	8
CBM15	PF03426.8	Carbohydrate binding domain (family 15)	0	0	0	0	3	0	0	5	4	4	1	4	1
CBM16	PF02018.11	Carbohydrate binding domain	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM17	PF03424.8	Carbohydrate binding domain (family 17/28)	0	0	0	0	0	0	0	1	1	0	1	1	0
CBM18	PF00187.13	Chitin recognition protein	4	2	5	3	1	0	0	38	35	26	13	9	5
CBM19	PF03427.7	Carbohydrate binding domain (family 19)	5	8	4	4	7	2	2	45	72	48	14	26	5
CBM20	PF00686.13	Starch binding domain	2	2	1	1	2	0	2	10	12	3	9	23	4
CBM21	PF03370.7	Putative phosphatase regulatory subunit	0	0	0	0	0	0	2	3	2	1	0	15	1
CBM22	PF02018.11	Carbohydrate binding domain	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM25	PF03423.7	Carbohydrate binding domain (family 25)	0	1	0	0	1	1	1	5	2	0	2	7	1
CBM27	PF09212.4	Carbohydrate binding module 27	0	0	0	1	0	0	0	1	0	0	0	3	19
CBM28	PF03424.8	Carbohydrate binding domain (family 17/28)	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM32	PF00754.19	F5/8 type C domain	2	5	5	7	13	0	0	51	49	24	34	97	12
CBM33	PF03067.9	Chitin binding domain	0	0	0	0	0	0	1	2	1	1	1	1	0
CBM34	PF02903.8	Alpha amylase, N-terminal ig-like domain	0	0	0	0	2	0	0	2	1	1	2	17	
CBM40	PF02973.10	Sialidase, N-terminal domain	5	3	1	5	0	0	0	23	24	8	11	28	2
CBM41	PF03714.8	Bacterial pullanase-associated domain	0	0	0	1	0	1	0	1	1	1	2	4	1
CBM42	PF05270.7	Alpha-L-arabinofuranosidase B (ABFB)	2	0	0	0	0	0	0	1	8	2	3	2	1
CBM43	PF07983.7	X8 domain	0	1	0	0	0	0	0	4	5	6	3	9	2
CBM47	PF00754.19	F5/8 type C domain	0	0	0	0	0	0	0	0	0	0	0		
CBM48	PF02922.12	Carbohydrate-binding module 48 (Isoamylase N-terminal domain)	5	3	3	4	7	0	3	66	51	45	23	76	7
CBM49	PF09478.4	Carbohydrate binding domain CBM49	0	0	2	0	0	1	0	3	6	4	1	19	2
CBM51	PF10633.3	NPCBM-associated, NEW3 domain of alpha-galactosidase	10	8	11	2	9	1	0	66	121	61	34	43	4

Other carbohydrate active enzymes			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
Other 1	PF07821.6	Alpha-amylase C-terminal beta-sheet domain	1	0	0	0	0	0	0	1	14	1	3	4	13
Other 2	PF02806.12	Alpha amylase, C-terminal all-beta domain	4	10	5	4	9	0	0	44	88	59	25	138	9
Other 3	PF09261.5	Alpha mannosidase, middle domain	0	0	1	2	0	0	0	6	9	2	6	12	1
Other 4	PF01510.19	N-acetylmuramoyl-L-alanine amidase	2	3	6	2	3	1	0	26	46	27	21	105	19
Other 5	PF01520.12	N-acetylmuramoyl-L-alanine amidase	6	3	4	2	10	1	0	66	66	63	35	222	15
Other 6	PF12123.2	N-acetylmuramoyl-l-alanine amidase	2	1	3	0	0	0	0	11	7	1	1	12	1
Other 7	PF08531.4	Alpha-L-rhamnosidase N-terminal domain	4	3	0	3	3	0	0	23	12	8	7	56	7
Other 8	PF13004.1	Bacteroidetes-Associated Carbohydrate-binding Often N-terminal	2	4	2	1	1	1	0	27	32	14	10	11	2
Other 9	PF10435.3	Beta-galactosidase, domain 2	0	0	0	0	0	0	0	1	0	1	1	3	1
Other 10	PF00144.18	Beta-lactamase	24	41	33	27	31	2	1	404	433	292	186	1599	140

	PF ID	Description													
Other 11	PF02929.11	Beta galactosidase small chain	0	3	2	2	8	0	0	5	8	5	4	8	2
Other 12	PF02369.10	Bacterial Ig-like domain (group 1)	6	9	11	5	18	1	0	134	132	75	48	135	9
Other 13	PF02368.12	Bacterial Ig-like domain (group 2)	2	3	7	3	6	0	3	92	76	35	31	45	3
Other 14	PF07523.6	Bacterial Ig-like domain (group 3)	1	2	1	1	4	0	0	12	19	10	10	69	3
Other 15	PF07532.5	Bacterial Ig-like domain (group 4)	0	0	1	0	0	0	0	7	10	4	2	21	0
Other 16	PF06204.5	Putative carbohydrate binding domain	0	4	1	3	7	0	0	21	31	25	14	18	5
Other 17	PF02927.8	N-terminal ig-like domain of cellulase	0	2	1	1	2	0	0	10	4	2	5	35	2
Other 18	PF12876.1	Sugar-binding cellulase-like	2	1	1	3	4	0	0	22	11	10	18	50	5
Other 19	PF03173.7	Putative carbohydrate binding domain	0	0	0	0	0	0	0	2	2	0	1	2	1
Other 20	PF03174.7	Chitobiase/beta-hexosaminidase C-terminal domain	1	1	4	1	3	0	0	26	18	19	13	84	17
Other 21	PF06483.5	Chitinase C	0	0	1	1	0	0	0	4	0	1	0	6	0
Other 22	PF08329.4	Chitinase A, N-terminal domain	0	1	1	0	2	0	0	13	9	11	8	33	5
Other 23	PF00963.12	Cohesin domain	1	2	2	1	1	0	0	20	24	6	13	9	0
Other 24	PF10438.3	Cyclo-malto-dextrinase C-terminal domain	0	0	1	1	1	0	0	4	8	3	2	50	7
Other 25	PF09087.5	Cyclomaltodextrinase, N-terminal	0	0	0	0	1	0	0	2	1	1	4	24	4
Other 26	PF00404.12	Dockerin type I repeat	6	9	3	4	6	0	0	34	55	21	17	44	2
Other 27	PF05838.6	Predicted lysozyme (DUF847)	0	0	0	0	0	0	0	0	0	0			
Other 28	PF07554.7	Uncharacterised Sugar-binding Domain	0	0	0	3	0	0	0	0	4	3	1	3	0
Other 29	PF00041.15	Fibronectin type III domain	10	22	11	12	26	1	0	114	130	67	91	216	22
Other 30	PF06202.8	Amylo-alpha-1,6-glucosidase	0	3	9	4	2	0	0	55	43	31	21	23	0
Other 31	PF09136.4	Glucodextranase, domain B	1	2	2	2	2	0	0	3	8	8	6	6	1
Other 32	PF09137.5	Glucodextranase, domain N	0	2	0	0	1	0	0	9	14	9	2	4	3
Other 33	PF02836.11	Glycosyl hydrolases family 2, TIM barrel domain	5	6	3	4	21	0	0	62	17	4	19	123	10
Other 34	PF02837.12	Glycosyl hydrolases family 2, sugar binding domain	2	10	1	6	18	0	0	58	53	13	20	158	13
Other 35	PF01915.16	Glycosyl hydrolase family 3 C terminal domain	13	8	6	12	18	0	0	78	45	28	40	115	17
Other 36	PF11975.1	Family 4 glycosyl hydrolase C-terminal domain	2	6	2	6	1	0	0	14	15	10	13	14	2
Other 37	PF02838.9	Glycosyl hydrolase family 20, domain 2	0	0	1	0	2	0	0	6	3	2	3	46	5
Other 38	PF08244.6	Glycosyl hydrolases family 32 C terminal	3	2	0	3	3	0	0	6	8	2	2	51	4
Other 39	PF07748.7	Glycosyl hydrolases family 38 C-terminal domain	1	3	1	4	8	0	0	20	9	3	9	12	2
Other 40	PF08533.4	Beta-galactosidase C-terminal domain	1	0	1	2	7	1	0	14	18	7	11	31	5
Other 41	PF08532.4	Beta-galactosidase trimerisation domain	1	4	3	2	5	0	0	34	27	5	18	25	0
Other 42	PF03633.9	Glycosyl hydrolase family 65, C-terminal domain	3	2	3	5	6	0	0	22	31	13	8	35	6
Other 43	PF03636.9	Glycosyl hydrolase family 65, N-terminal domain	0	0	0	1	7	0	0	16	9	5	1	74	2
Other 44	PF07477.6	Glycosyl hydrolase family 67 C-terminus	1	1	0	1	1	0	0	10	3	0	10	27	3
Other 45	PF03648.8	Glycosyl hydrolase family 67 N-terminus	0	0	0	2	1	1	0	10	3	2	4	36	6
Other 46	PF08307.5	Glycosyl hydrolase family 98 C-terminal domain	0	0	0	0	0	0	0	1	1	0	0	5	1
Other 47	PF11790.2	Glycosyl hydrolase catalytic core	1	1	1	0	1	0	1	9	17	7	12	19	1
Other 48	PF00852.13	Glycosyltransferase family 10 (fucosyltransferase)	0	0	0	0	0	0	0	5	11	2	2	42	0
Other 49	PF01755.11	Glycosyltransferase family 25 (LPS biosynthesis protein)	1	2	0	0	0	0	0	8	7	2	4	19	2
Other 50	PF06165.5	Glycosyltransferase family 36	2	4	6	2	7	0	0	10	18	9	17	12	6
Other 51	PF00534.14	Glycosyl transferases group 1	70	85	75	94	141	1	0	605	1015	617	532	1678	148
Other 52	PF00535.20	Glycosyl transferase family 2	23	37	32	36	66	1	1	531	381	323	360	1926	114
Other 53	PF06205.5	Glycosyltransferase 36 associated family	1	5	8	6	7	0	0	24	47	30	39	5	2
Other 54	PF05345.6	Putative Ig domain	9	12	6	6	8	1	0	104	142	72	29	97	22
Other 55	PF00206.1	Lyase	14	6	10	7	8	1	0	166	113	102	64	370	29
Other 56	PF02884.11	Polysaccharide lyase family 8, C-terminal beta-sandwich domain	0	1	1	0	0	0	0	1	3	5	2	17	0
Other 57	PF08124.5	Polysaccharide lyase family 8, N terminal alpha-helical domain	0	0	0	0	0	0	0	5	5	2	0	14	1
Other 58	PF07555.7	beta-N-acetylglucosaminidase	0	0	0	0	0	0	0	2	4	3	2	7	2
Other 59	PF12972.1	Alpha-N-acetylglucosaminidase (NAGLU) C-terminal domain	0	0	0	0	0	0	0	8	6	1	2	35	4
Other 60	PF12971.1	Alpha-N-acetylglucosaminidase (NAGLU) N-terminal	2	1	0	1	0	0	0	12	9	4	2	12	5

			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
		domain													
Other 61	PF01471.12	Putative peptidoglycan binding domain	7	15	7	7	8	0	1	77	175	187	57	98	16
Other 62	PF02027.11	RolB/RolC glucosidase family	1	0	2	0	0	0	0	4	6	3	4	6	2
Other 63	PF06283.5	Trehalose utilisation	7	2	11	6	11	1	2	78	177	67	43	236	13
Other 64	PF01833.18	IPT/TIG domain	8	5	8	3	14	1	0	103	86	65	34	64	5
Other 65	PF07492.5	Neutral trehalase Ca2+ binding domain	0	0	0	0	0	0	0	4	0	4	1	5	0
Other 66	PF06955.6	Xyloglucan endo-transglycosylase (XET) C-terminus	1	0	2	0	0	0	0	5	5	1	1	5	1
Phenolic compound degradation domains			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
Laccase	PF02578.9		1	8	7	11	12	1	0	34	49	49	18	53	5
Dioxygenase	PF00775.15		2	8	3	2	5	0	0	51	85	35	22	34	6
Peroxidase	PF00141.17		4	5	3	8	1	0	0	23	36	17	4	83	4
Polysaccharide lyase			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
PL3	PF12708.1	Pectate lyase superfamily protein	2	5	3	7	8	0	0	50	29	21	29	68	8
PL5	PF05426.6		0	0	1	1	1	0	0	14	10	16	8	38	4
PL8	PF02278.12	Polysaccharide lyase family 8, super-sandwich domain	0	0	0	0	0	0	0	1	3	0	1	25	2
Carbohydrate esterase			S2a	S2b	K1a	K1b	K1c	Solvatn mRNA	Knudsenheia mRNA	Luquillo	Rotham 0-10	Rotham 10-21	Waseca	Eureka top	Eureka 2m
CE4	PF01522.15	Polysaccharide deacetylase	10	8	9	15	21	1	0	170	109	107	74	282	31
CE5	PF05448		7	9	6	10	13	1	0	115	116	75	94	258	20
CE8	PF01095.11		3	2	0	0	4	0	0	11	10	7	2	60	3
CE11	PF03331.5		0	0	0	0	0	0	0	0	0	0	0	0	0
CE11 (2)	PF03331.7		10	3	1	5	6	0	0	35	62	19	13	112	8
CE14	PF02585		3	7	6	5	6	1	0	74	57	37	61	183	21

Table S5. Description of soil metagenomes included in comparative analyses

Name	Accession	Description	Method	Reference
Luquillo experimental forest, Puerto Rico	SRX015858	The top 10cm of a subtropical wet forest soil	454-Random whole genome shotgun library	-
Rothamsted, UK	454-0709-0-10.tar.gz (https://metasoil.univlyon1.fr)	The top 0-10cm of a grass field soil	454-Random whole genome shotgun library	-
Rothamsted, UK	454-0309-10-21.tar.gz (https://metasoil.univlyon1.fr)	The lower 10-21cm of a grass field soil	454-Random whole genome shotgun library	-
Waseca, Minnesota, USA	AAFX00000000	The top 10cm of a farmland soil	454-Random whole genome shotgun library	Tringe, von Mering et al. 2005
Eureka, Nunavut, Canada	SRX016196	The active layer of a permafrost mineral soil, sedges, other vascular plants, and mosses	454-Random whole genome shotgun library	Yergeau, Hogues et al. 2010
Eureka, Nunavut, Canada	SRX016196	The 2m depth permafrost layer of a permafrost mineral soil	454-Random whole genome shotgun library	Yergeau, Hogues et al. 2010

Table S6. PFAMs that target plant structural polymers, identified in metagenomes from soil metagenomes
Numbers reflect the number of significant hits in the metagenomic libraries generated from the Svalbard peat soils. Shown in brackets are the numbers in $^0/_{000}$, of the total number of metagenomic sequences assigned to a protein family domain. n.d. implies that the respective protein family was not detected.

GH	Function	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted _0-10cm	Rothamsted_10-21cm	Waseca	Eureka active layer	Eureka permafrost
Cellulases												
GH5	cellulase	4 (2.3)	4 (1.8)	4 (2.1)	8 (3.9)	18 (5.8)	35 (1.9)	56 (2.3)	22 (1.6)	35 (4.3)	144 (4.4)	13 (2.5)
GH6	endoglucanase	2 (1.2)	0 (0)	1 (0.5)	1 (0.5)	2 (0.7)	2 (0.1)	8 (0.3)	3 (0.2)	5 (0.6)	3 (0.1)	1 (0.2)
GH7	endoglucanase	0 (0)	1 (0.5)	0 (0)	1 (0.5)	0 (0)	3 (0.2)	4 (0.2)	3 (0.2)	3 (0.4)	7 (0.2)	2 (0.4)
GH9	endoglucanase	1 (0.5)	2 (0.9)	4 (2.1)	1 (0.5)	6 (2.0)	26 (1.4)	9 (0.4)	3 (0.2)	13 (1.6)	92 (2.8)	8 (1.5)
GH44	endoglucanase	2 (1.2)	2 (0.9)	3 (1.6)	2 (1.0)	0 (0)	9 (0.5)	12 (0.5)	3 (0.2)	2 (0.2)	5 (0.2)	0 (0)
GH45	endoglucanase	1 (0.6)	0 (0)	0 (0)	0 (0)	1 (0.3)	5 (0.3)	12 (0.5)	4 (0.3)	1 (0.1)	8 (0.2)	1 (0.2)
GH48	endo-processive cellulases	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.1)	3 (0.1)	1 (0.1)	1 (0.1)	5 (0.2)	0 (0)
Sum		10 (5.9)	9 (4.1)	12 (6.3)	13 (6.3)	27 (8.8)	91 (5.0)	104 (4.3)	39 (2.8)	60 (7.3)	264 (8.0)	25 (4.8)
Debranching enzymes												
GH51	a-L-arabinofuranosidase	5 (2.9)	8 (3.6)	3 (1.6)	5 (2.4)	13 (4.2)	33 (1.8)	30 (1.2)	8 (0.6)	12 (1.5)	19 (0.6)	6 (1.2)
GH54	a-L-arabinofuranosidase	1 (0.6)	3 (1.4)	0 (0)	0 (0)	0 (0)	5 (0.3)	9 (0.4)	0 (0)	2 (0.2)	4 (0.1)	0 (0)
GH62	a-L-arabinofuranosidase	0 (0)	0 (0)	1 (0.5)	0 (0)	0 (0)	2 (0.1)	6 (0.2)	3 (0.2)	3 (0.4)	8 (0.2)	0 (0)
GH67	a-glucuronidase	2 (1.2)	1 (0.5)	0 (0)	0 (0)	2 (0.7)	16 (0.9)	3 (0.1)	3 (0.2)	3 (0.4)	18 (0.5)	5 (1.0)
GH78	a-L-rhamnosidase	1 (0.6)	2 (0.9)	2 (1.1)	10 (4.9)	15 (4.9)	55 (3.0)	24 (1.0)	10 (0.7)	15 (1.8)	154 (4.7)	18 (3.5)
Sum		9 (5.3)	14 (6.4)	6 (3.1)	15 (7.3)	30 (9.7)	124 (6.9)	72 (3.0)	25 (1.8)	35 (4.3)	203 (6.2)	29 (5.6)
Endohemicellulases												
GH8	endo-xylanases	0 (0)	0 (0)	1 (0.5)	0 (0)	2 (0.7)	6 (0.3)	7 (0.3)	5 (0.4)	4 (0.5)	8 (0.2)	1 (0.2)
GH10	endo-1,4-b-xylanase	2 (1.2)	6 (2.7)	1 (0.5)	5 (2.4)	7 (2.3)	19 (1.1)	28 (1.2)	18 (1.3)	19 (2.3)	68 (2.1)	14 (2.7)
GH11	xylanase	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.3)	8 (0.4)	5 (0.2)	5 (0.4)	1 (0.1)	7 (0.2)	1 (0.2)
GH12	endoglucanase & xyloglucan hydrolysis	0 (0)	1 (0.5)	0 (0)	0 (0)	0 (0)	2 (0.1)	2 (0.1)	1 (0.1)	4 (0.5)	2 (0.1)	0 (0)
GH26	b-mannanase & xylanase	4 (2.3)	4 (1.8)	0 (0)	1 (0.5)	12 (3.9)	7 (0.4)	9 (0.4)	9 (0.6)	5 (0.6)	34 (1.0)	2 (0.4)
GH28	galacturonases	0 (0)	1 (0.5)	4 (2.1)	2 (1.0)	8 (2.6)	20 (1.1)	10 (0.4)	4 (0.3)	8 (1.0)	62 (1.9)	6 (1.2)
GH53	endo-1,4-b-galactanase	0 (0)	3 (1.4)	3 (1.6)	2 (1.0)	4 (1.3)	13 (0.7)	10 (0.4)	0 (0)	6 (0.7)	28 (0.9)	5 (1.0)
Sum		6 (3.5)	15 (6.8)	9 (4.7)	10 (4.9)	34 (11.0)	80 (4.4)	71 (3.0)	42 (3.0)	47 (5.7)	209 (6.4)	29 (5.6)
Oligosaccharide degrading enzymes												
GH1	b-glucosidase and many other b-linked dimers	9 (5.3)	10 (4.6)	6 (3.1)	14 (6.8)	20 (6.5)	79 (4.4)	53 (2.2)	41 (2.9)	24 (2.9)	65 (2.0)	11 (2.1)
GH2	b-galactosidases and other b-linked dimers	5 (2.9)	0 (0)	4 (2.1)	4 (1.9)	15 (4.9)	18 (1.0)	6 (0.2)	6 (0.4)	11 (1.3)	72 (2.2)	4 (0.8)
GH3	mainly b-glucosidases	5 (2.9)	9 (4.1)	6 (3.1)	9 (4.4)	18 (5.8)	94 (5.2)	53 (2.2)	38 (2.7)	42 (5.1)	304 (9.3)	28 (5.4)
GH29	a-L-fucosidase	5 (2.9)	4 (1.8)	4 (2.1)	11 (5.3)	13 (4.2)	58 (3.2)	29 (1.2)	4 (0.3)	23 (2.8)	62 (1.9)	21 (4.1)
GH35	b-galactosidase	2 (1.2)	0 (0)	0 (0)	0 (0)	3 (1.0)	15 (0.8)	16 (0.7)	2 (0.1)	5 (0.6)	4 (0.1)	4 (0.8)
GH38	a-mannosidase	0 (0)	0 (0)	1 (0.5)	3 (1.5)	4 (1.3)	27 (1.5)	7 (0.3)	7 (0.5)	9 (1.1)	16 (0.5)	3 (0.6)
GH39	b-xylosidase	2 (1.2)	5 (2.3)	1 (0.5)	2 (1.0)	7 (2.3)	21 (1.2)	12 (0.5)	7 (0.5)	15 (1.8)	22 (0.7)	3 (0.6)
GH42	b-galactosidase	3 (1.8)	2 (0.9)	2 (1.1)	6 (2.9)	11 (3.6)	47 (2.6)	27 (1.1)	14 (1.0)	23 (2.8)	46 (1.4)	8 (1.5)
GH43	arabinases & xylosidases	8 (4.7)	7 (3.2)	5 (2.6)	8 (3.9)	10 (3.2)	59 (3.3)	50 (2.1)	12 (0.9)	46 (5.6)	372 (11.3)	66 (12.8)
GH52	b-xylosidase	0 (0)	0 (0)	0 (0)	2 (1.0)	0 (0)	3 (0.2)	0 (0)	1 (0.1)	1 (0.1)	1 (0.03)	0 (0)
Sum		39 (22.9)	37 (16.8)	29 (15.2)	59 (28.7)	101 (32.8)	421 (23.3)	257 (10.7)	136 (9.7)	199 (24.3)	964 (29.4)	148 (28.6)
Phenolic compound degrading enzymes												
Laccase		1 (0.6)	8 (3.6)	7 (3.7)	11 (5.3)	12 (3.9)	32 (1.8)	47 (2.0)	46 (3.3)	18 (2.2)	16 (0.5)	4 (0.8)
Dioxygenase		2 (1.2)	8 (3.6)	3 (1.6)	2 (1.0)	5 (1.6)	51 (2.8)	85 (3.5)	35 (2.5)	22 (2.7)	34 (1.0)	6 (1.2)
Peroxidase		4 (2.3)	5 (2.3)	3 (1.6)	8 (3.9)	1 (0.3)	23 (1.3)	36 (1.5)	17 (1.2)	4 (0.5)	83 (2.5)	4 (0.8)
Sum		7 (4.1)	21 (9.6)	13 (6.8)	21 (10.2)	18 (5.8)	106 (5.9)	168 (7.0)	98 (7.0)	48 (5.4)	133 (4.1)	14 (2.7)
Total assigned pfam		17067	21980	19104	20581	30834	180356	240284	140447	81750	328389	51760

Table S7. Chi-square test statistics from pairwise comparison of the fraction of genes encoding plant polymer degrading enzymes in soil metagenomes. The fraction of reads assigned to each of five categories of plant polymer degradation (Table S6) was compared between each of the soil metagenomes (Table S5). The table shows the evaluated chi-square test statistic values for each comparison. Colored cells indicate three different levels of confidence (p-values 0.05, 0.01, 0.001). In the cases where the cells are colored, the two compared soil metagenomes contain a significantly different number of sequences of the given category relative to the total number of assignable sequences (all pfam's).

P-value	0.05	0.01	0.001							

Cellulases	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted_0-10cm	Rothamsted_10-21cm	Waseca	Eureka active layer
S2a										
S2b	0.62									
K1a	0.03	0.96								
K1b	0.03	1.02	0							
K1c	1.19	4.09	0.93	0.95						
Luquillo	0.58	0.09	1.1	1.22	9.05					
Rothamsted_0_10cm	0.84	0.03	1.51	1.67	11.1	0.11				
Rothamsted_10_21cm	4.65	1.12	6.46	6.97	23.47	6.56	5.68			
Waseca	0.44	2.74	0.24	0.24	0.58	8.1	10.87	24.15		
Eureka active layer	0.97	4.12	0.7	0.72	0.18	20.9	29.55	42.18	0.41	
Eureka permafrost	0.27	0.18	0.56	0.62	4.74	0.07	0.24	4.79	3.14	6.06

Debranching enzymes	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted_0_10cm	Rothamsted_10_21cm	Waseca	Eureka active layer
S2a										
S2b	0.2									
K1a	0.99	2.19								
K1b	0.59	0.13	3.22							
K1c	2.68	1.74	7.11	0.84						
Luquillo	0.2	0.01	2.68	0.38	5.04					
Rothamsted_0_10cm	2.63	6.99	0.01	10.47	32.94	23.63				
Rothamsted_10_21cm	9.23	17.65	1.83	23.08	51.61	37.1	5.83			
Waseca	0.31	1.6	0.5	3.04	11.52	3.55	3.03	12.88		
Eureka active layer	0.97	4.12	0.7	0.72	0.18	20.9	29.55	42.18	0.41	
Eureka permafrost	0.27	0.18	0.56	0.62	4.74	0.07	0.24	4.79	3.14	6.06

Endohemicellulases	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted_0_10cm	Rothamsted_10_21cm	Waseca	Eureka active layer
S2a										
S2b	1.96									
K1a	0.31	0.78								
K1b	0.4	0.7	0							
K1c	7.43	2.44	5.47	5.49						
Luquillo	0.16	3.13	0.13	0.21	24.08					
Rothamsted_0_10cm	0.17	9.2	1.77	2.21	45.99	4.3				
Rothamsted_10_21cm	0.14	7.96	1.56	1.94	36.82	2.95	0			
Waseca	1.31	0.34	0.3	0.23	8.67	3.06	13	9.82		
Eureka active layer	2.12	0.07	0.79	0.7	9.06	10.15	32.78	20.93	0.4	
Eureka permafrost	1.1	0.39	0.21	0.15	7.46	1.87	8.72	6.99	0.01	0.41

Oligosaccharide degrading enzymes	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted_0_10cm	Rothamsted_10_21cm	Waseca	Eureka active layer
S2a										
S2b	1.79									
K1a	2.83	0.17								
K1b	1.22	6.61	8.15							
K1c	3.7	12.48	14.03	0.67						
Luquillo	0.02	3.68	5.11	2.2	9.46					
Rothamsted_0_10cm	21.35	7.25	3.52	52.21	103.53	105.74				
Rothamsted_10_21cm	25.4	10.11	5.58	56.26	101.54	89.22	1.12			
Waseca	0.13	4.3	5.76	1.22	5.96	0.24	83.04	77.58		
Eureka active layer	2.37	11.34	12.73	0.03	1.1	15.5	230.29	167.99	5.82	

| Eureka permafrost | 1.56 | 8.53 | 10.08 | 0 | 1.11 | 4.53 | 101.35 | 95.8 | 2.21 | 0.09 |

Phenolic compound degrading enzymes	S2a	S2b	K1a	K1b	K1c	Luquillo	Rothamsted_0_10cm	Rothamsted_10_21cm	Waseca	Eureka active layer
S2a										
S2b	3.99									
K1a	1.19	0.93								
K1b	2.53	0.2	0.28							
K1c	0.44	2.91	0.33	1.41						
Luquillo	0.64	5.08	0.44	2.23	0					
Rothamsted_0_10cm	1.4	2.89	0.03	0.94	0.38	1.23				
Rothamsted_10_21cm	1.18	3.06	0.08	1.1	0.23	0.59	0.05			
Waseca	0.24	5.97	0.93	3.06	0.11	0.36	2.07	1.39		
Eureka active layer	0.02	15.51	3.76	8.87	1.82	7.5	18.18	11.98	1.98	
Eureka permafrost	0.82	15.26	6.16	10.61	4.06	6.81	10.33	9.04	4.11	1.7