LETTER

doi:10.1038/nature10576

Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw

Rachel Mackelprang[1,2], Mark P. Waldrop[3], Kristen M. DeAngelis[4], Maude M. David[4], Krystle L. Chavarria[4], Steven J. Blazewicz[5], Edward M. Rubin[2,6] & Janet K. Jansson[2,4]

Permafrost contains an estimated 1672 Pg carbon (C), an amount roughly equivalent to the total currently contained within land plants and the atmosphere[1–3]. This reservoir of C is vulnerable to decomposition as rising global temperatures cause the permafrost to thaw[2]. During thaw, trapped organic matter may become more accessible for microbial degradation and result in greenhouse gas emissions[4,5]. Despite recent advances in the use of molecular tools to study permafrost microbial communities[6–9], their response to thaw remains unclear. Here we use deep metagenomic sequencing to determine the impact of thaw on microbial phylogenetic and functional genes, and relate these data to measurements of methane emissions. Metagenomics, the direct sequencing of DNA from the environment, allows the examination of whole biochemical pathways and associated processes, as opposed to individual pieces of the metabolic puzzle. Our metagenome analyses reveal that during transition from a frozen to a thawed state there are rapid shifts in many microbial, phylogenetic and functional gene abundances and pathways. After one week of incubation at 5 °C, permafrost metagenomes converge to be more similar to each other than while they are frozen. We find that multiple genes involved in cycling of C and nitrogen shift rapidly during thaw. We also construct the first draft genome from a complex soil metagenome, which corresponds to a novel methanogen. Methane previously accumulated in permafrost is released during thaw and subsequently consumed by methanotrophic bacteria. Together these data point towards the importance of rapid cycling of methane and nitrogen in thawing permafrost.

We collected three intact frozen permafrost soil cores with their overlying seasonally thawed active layers from Hess Creek, Alaska. This is a black-spruce forest site containing many metres of frozen peat and the C was dated to 1200 yr BP[10]. Other soil properties and microbial respiration rates were previously characterized[10]. Frozen active layer and permafrost layer samples from each core were thawed and incubated for 7 days at 5 °C under a helium headspace. During the incubations, CH_4 (Fig. 1a) and CO_2 (Supplementary Fig. 1) concentrations were monitored in the headspace and DNA was extracted for 16S ribosomal RNA (rRNA) and metagenome sequencing.

There was a burst of CH_4 from the permafrost within 48 h of thaw, followed by a significant ($P = 0.05$) decrease in concentration from day 2 to day 7 (Fig. 1a). To determine (1) if methane release was due to post-thaw production or from trapped gas, (2) whether the methane was consumed by methanotrophs or anaerobic methane oxidizers, and (3) the CH_4 oxidation potential over time, we treated additional samples with 1500 p.p.m. CH_4 and 2-bromoethane sulphonic acid (BES) and measured CH_4 levels daily. BES is an inhibitor of archaeal methanogenesis and methyl-coenzyme M reductase-dependent anaerobic methane oxidation. Rapid release of CH_4 from samples treated with BES suggested that the CH_4 primarily originated from gas present in the permafrost before thaw (Supplementary Fig. 2), as previously reported[11]. Subsequent CH_4 consumption in both BES- and non-BES-treated samples was indicative of CH_4 oxidation by methanotrophic bacteria (Fig. 1b). The oxygen used for methane oxidation presumably originated from permafrost water or aerobic microsites in the samples[12]. Together these data indicate CH_4 levels are dynamic in thawing permafrost.

To determine the phylogenetic and functional gene repertoire before and after thaw, we performed deep metagenome sequencing of samples from two of the three replicate cores (cores 1 and 2). DNA was extracted from frozen active layer and permafrost samples and from samples thawed at 5 °C for 2 and 7 days. This resulted in 12 samples for metagenome sequencing.

Owing to low DNA yield we used emulsion PCR (emPCR) to generate random shotgun short insert libraries with minimum amplification bias[13]. Sequencing yielded a total of 176 million reads and 39.8 Gb of raw sequence. The individual metagenome reads were annotated by comparison with protein-coding and small subunit (SSU) rRNA genes (Supplementary Table 1). We achieved relatively good assembly of the metagenome data, despite the high microbial diversity of soil. In total we obtained 9.7 Mb of sequence in 3,758 contigs greater than 1 kb in length. The longest contig was 67.4 kb in length and 98 contigs were greater than 10 kb.

A draft genome of a novel methanogen was constructed from the metagenome data (Fig. 2), owing to the relatively high abundance of methanogens in the samples (0.2–4% of total 16S rRNA gene sequences) and its low population heterogeneity. To generate the draft genome, contigs were binned by tetranucleotide frequency and read coverage[14]. The draft genome comprised 1.9 Mb of sequence in 174 contigs. To confirm that the clustered contigs represented a draft genome, we validated the assembly through single-copy gene analysis (Supplementary Table 2) and by alignment of the draft genome to a related reference genome (Supplementary Fig. 3). This is the first example of successful assembly of a draft genome from a highly complex soil metagenome. The nearest sequenced relative was the recently

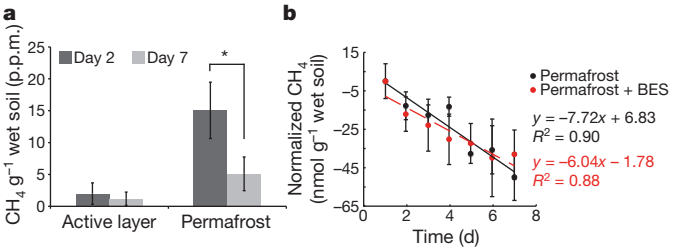

Figure 1 | Bacterial oxidation of trapped CH_4 released from Hess Creek permafrost after thaw. **a**, CH_4 headspace concentration during incubation of permafrost and active layer samples at 5 °C for 2 and 7 days. Values are means ± s.e.m. **b**, CH_4 headspace concentrations during 7 days incubation at 5 °C from permafrost samples without BES and with BES. Values are normalized to CH_4 concentrations after the initial release of CH_4 during thaw (day 1). *$P = 0.05$ ($n = 3$).

[1]Department of Biology, California State University at Northridge, Northridge, California 91330, USA. [2]United States Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. [3]United States Geological Survey, Menlo Park, California 94025, USA. [4]Lawrence Berkeley National Laboratory, Earth Sciences Division, Berkeley, California 94720, USA. [5]Department of Environmental Science, Policy and Management, University of California at Berkeley, Berkeley, California 94720, USA. [6]Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, California 94720, USA.

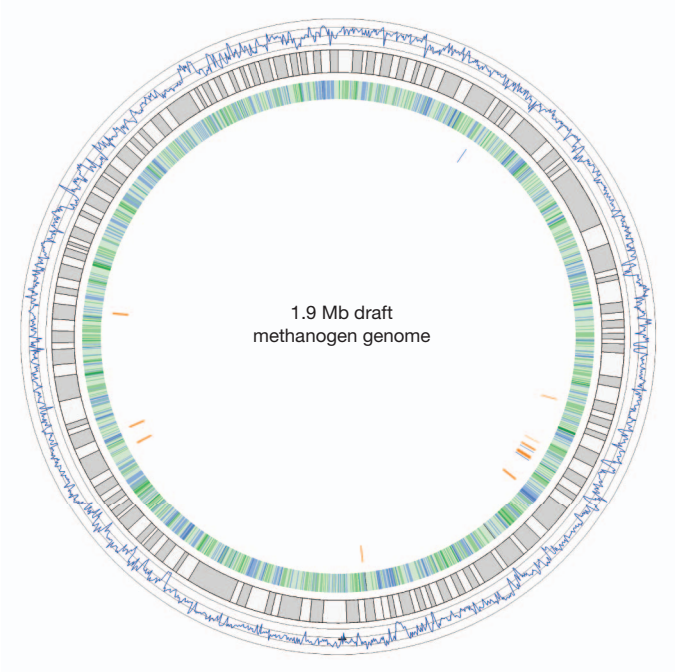

Figure 2 | Draft methanogen genome assembly. Draft methanogen genome. Features correspond to concentric circles, starting with the outermost circle. (1) Illumina sequence coverage averaging 73×. (2) One hundred and seventy-four contigs making up the draft genome; contigs shown are scaled according to size and are roughly ordered by mapping to the reference genome (65% identity at the nucleotide level). (3) GC content heat map (dark blue to light green represent low to high values). (4) Methanogenesis genes (orange) and nitrogen fixation genes (blue). The true size of the genome is not known owing to gaps between the contigs.

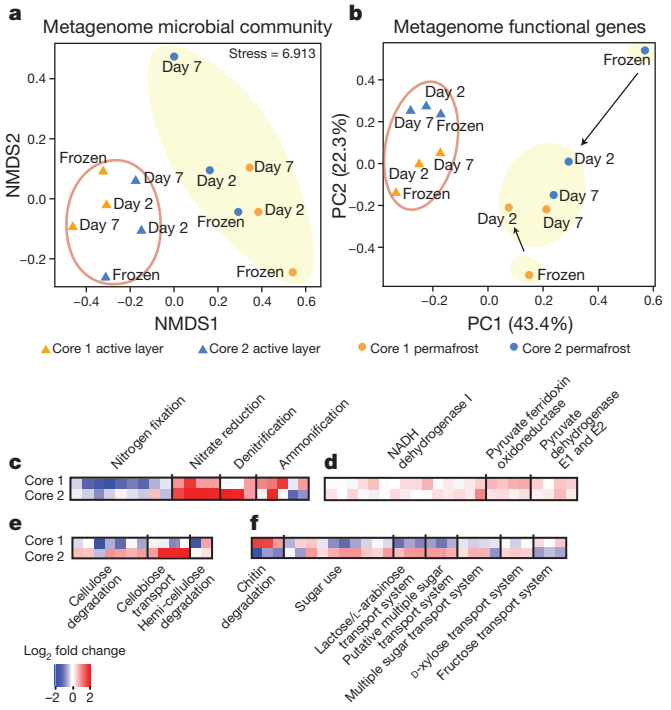

Figure 3 | Thaw-induced shifts of phylogenetic and functional genes in metagenomes. **a**, nMDS analysis of the relative abundance of 16S rRNA genes from the metagenomes. **b**, Principal component analysis of relative abundance of KEGG genes in metagenomes. The percentage variation explained by the principal components is indicated on the axes. Arrows illustrate rapid shift in functional gene composition upon thaw in two disparate permafrost samples. **c**–**f**, Heat maps indicating differences in relative abundances of specific genes between frozen (day 0) and thawed (day 7) permafrost metagenomes (Hess Creek cores 1 and 2). **c**, Nitrogen cycle; **d**, central metabolism; **e**, cellulose degradation; **f**, chitin degradation, sugar metabolism and transport.

described Methanocellales[15] order at a nucleotide identity level of approximately 65% (Fig. 2). Single-copy gene analysis demonstrated it was related to members of *Methanomicrobia* (Supplementary Fig. 4). The abundance of this novel methanogen correlates with the observed CH_4 in the samples and suggests that it may be an important player in CH_4 production under frozen conditions. It has previously been reported that trapped CH_4 in permafrost is biological in origin and that methanogenesis can occur at sub-zero temperatures[11]. The draft genome also included genes for nitrogen fixation. Although nitrogen-fixing methanogens have been previously described[16], this draft genome is the first indication that they are present in permafrost.

The metagenome data revealed core-specific shifts in some community members (Fig. 3a), including the orders Proteobacteria, Bacteriodetes and Firmicutes. We found that Actinobacteria increased in both cores during thaw (Supplementary Fig. 5). Actinobacteria have previously been found at high abundance in permafrost[9], which is thought to be caused by their maintenance of metabolic activity and DNA repair mechanisms at low temperatures[17]. Most archaeal sequences identified in the metagenomic data were methanogens in the phylum Euryarchaeota (62–95%), including the Methanomicrobia that was represented in our draft genome. In total, four orders of methanogens (Methanosarcinales, Methanomicrobiales, Methanomicrobia and Methanobacterales) were detected. As the permafrost thawed, the methanogens (including Methanomicrobia) increased in relative abundance (Supplementary Fig. 6). These orders are known to be metabolically versatile and can use a variety of substrates[18].

18S rRNA gene sequences from land plants (Streptophyta) were the most abundant eukaryotic reads in the metagenome data, probably originating from undecomposed detritus. 18S rRNA gene sequences also originated from fungi, protists, ameobae, algae and other eukaryotic phyla (Supplementary Fig. 7.) Few consistent changes in the Eukarya were observed after thaw, although the Streptophyta decreased in core 2, presumably owing to microbial degradation of plant material (Supplementary Fig. 7).

A greater phylogenetic distance was observed between frozen and day 2 samples than between day 2 and day 7 samples (Supplementary Fig. 8), based on 454 pyrotag sequencing of 16S rRNA genes, suggesting that the community composition shifted rapidly upon thaw. The difference was more pronounced in the permafrost than in the active layer. Operational taxonomic units changing significantly ($P < 0.05$) in abundance during thaw were largely from uncultivated taxa (Supplementary Fig. 9 and Supplementary Table 3).

We used quantitative PCR (qPCR) to measure the absolute abundances of specific phyla before and after thaw. The qPCR results confirmed that there was a significant increase in Actinobacteria in both cores after thaw, Bacteriodetes changed in a core-dependent manner, and no significant changes were observed in Chloroflexi (Supplementary Fig. 10).

Our observation that methane was consumed after thaw (Fig. 1) was correlated to detection of sequences representative of bacterial methanotrophs in relatively high amounts (approximately 0.25–0.65% relative abundance). Two forms of methane monooxygenases were detected: particulate methane monooxygenase (*pmoA*) represented most (~80%) and the rest were soluble methane monooxygenase (*mmoX*). The metagenomic results were confirmed by qPCR of *pmoA*, *mcrA* (encoding the methyl coenzyme-M reductase alpha subunit) and 16S rRNA genes from type I and type II methanotrophs. Both the *pmoA* gene and type II methanotrophs significantly increased in abundance after thaw ($P < 0.01$). Although type I methanotrophs were detectable at low levels (fewer than 100 copies per nanogram), they did not differ in abundance between the frozen and thawed samples. *McrA* sequences from methanogenic archaea were detected but did not change significantly during thaw (Supplementary Fig. 11). *McrA* and 16S sequences

originating from anaerobic methane oxidizers were not detected. Collectively, these data and our BES incubation experiments (Fig. 1b) suggest that bacterial type II methanotrophs and the particulate methane monooxygenase enzyme were involved in consumption of methane released during permafrost thaw.

We tracked simultaneous shifts in the total gene complement from the metagenome data to obtain a global view of functional response to thaw. The active layer samples were relatively similar before and after thaw. By contrast, the two frozen permafrost metagenomes differed dramatically before thaw (Fig. 3b). In addition, functional genes in frozen active layer and permafrost samples were distinct from each other, including differences in several key metabolic pathways such as energy metabolism, nitrogen fixation, amino-acid transport, oxidative phosphorylation and anaerobic respiration (Supplementary Fig. 12). During thaw, the permafrost metagenomes rapidly converged and neared those in the active layer samples (Fig. 3b). The convergence of function was not matched by a convergence of phylogenetic composition during this short-term incubation (Fig. 3a), suggesting that disparate community responses to thaw can have similar functional consequences.

The high sequencing depth used provided the sensitivity necessary to characterize subtle changes in the abundance of thousands of genes that had a cumulative significant affect, as confirmed by multiple statistical analyses (Supplementary Figs 13–15 and Supplementary Table 4). We specifically targeted genes involved in C and N cycling by grouping Kyoto Encyclopedia of Gene and Genomes (KEGG) genes into subsets (Fig. 3c–f). Several genes involved in the N cycle shifted in abundance during thaw (Fig. 3c). For example, nitrate reductase I genes significantly increased, suggesting nitrate was available as a terminal electron acceptor, which was confirmed by its presence in the chemical data[10]. Ammonification and denitrification genes increased during thaw in both cores. However, core 1 did not show an increase in the *nosZ* gene, responsible for converting N_2O to N_2, suggesting that it had potential for N_2O greenhouse gas emission during thaw. By contrast, several genes in the nitrogen fixation pathway decreased during thaw in both cores. Changes in the N cycle observed in the metagenomic data were confirmed by qPCR of specific genes for nitrate reduction (*narG*, encoding nitrate reductase I, alpha subunit) and nitrogen fixation (*nifH*, encoding the nitrogenase iron protein), which increased and decreased in abundance after thaw, respectively (Supplementary Fig. 16).

The metagenomic strategy also allowed us to identify changes in a diverse set of C cycle genes not easily accessed by other approaches such as qPCR or analysis of enzymatic activities. These included genes involved in central metabolism that increased during thaw (Fig. 3d). Genes involved in the transport and metabolism of specific carbohydrates were also enriched after thaw, but these changes were mainly core specific. For example, in core 2, we observed significant increases ($P < 0.01$) in the cellobiose transport system and a corresponding increase in cellulose degradation and hemicellulose degradation (Fig. 3f). Chitin degradation genes and some sugar transport systems were specifically enriched in core 1 (Fig. 3g) (see Supplementary Table 5 for a complete list). It is important to keep in mind that decreases in relative abundance may also be due to growth of other community members lacking those particular genes. These data reflect a rapid response of specific members of the microbial community in permafrost to thaw, probably owing to the availability of particular substrates that become accessible. Although the cores were taken from the same site within metres of each other and would presumably have similar chemical and physical properties, there were subtle differences between the genetic responses to thaw in the replicate cores. For example, we observed a higher level of dissolved organic carbon and higher C density in core 2 (Supplementary Table 6) that is correlated with the core-2-specific enrichment of C processing genes during thaw. These findings emphasize the importance of replication, even for complex soil metagenomes.

In summary, these detailed analyses reveal for the first time the rapid and dynamic response of permafrost microbial communities to thaw. The thaw-induced shifts that we detected directly support conceptual models of C and N cycling in Arctic soils (Fig. 4), in which microbes play a central role in greenhouse gas emissions and destabilization of stored permafrost C.

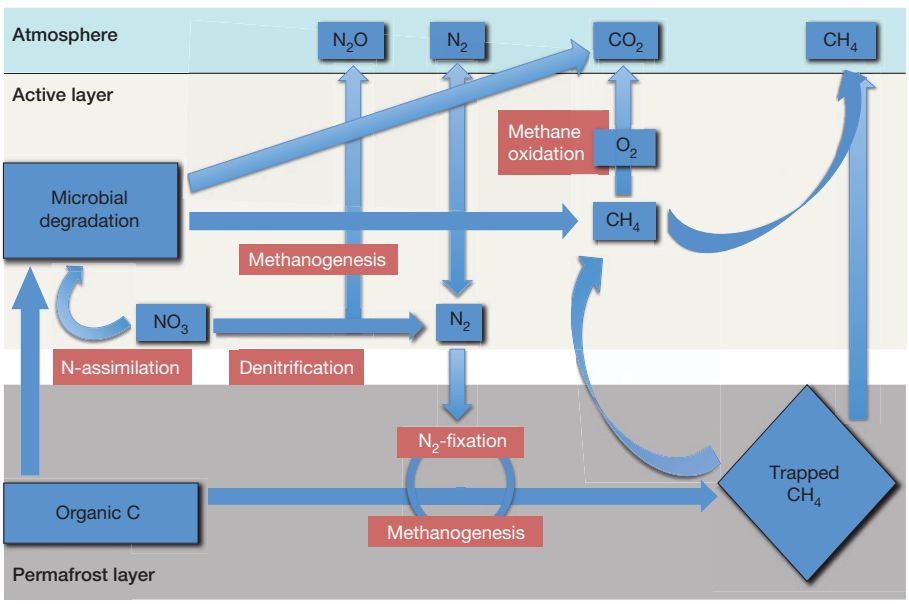

Figure 4 | Conceptual model of C and N cycling in Arctic soils based on metagenome data. Slow rates of methanogenesis by cold-adapted methanogens occur in permafrost and active layer soils. Over time, methane accumulates in the permafrost and is initially consumed by bacterial methanotrophs upon thaw. In the permafrost, N_2 fixation genes are abundant. As the permafrost thaws, microbial degradation of organic C occurs rapidly and other nitrogen sources become available with an increase in dissimilatory and assimilatory nitrate reduction processes.

METHODS SUMMARY

Triplicate frozen permafrost cores were collected from Hess Creek, Alaska, to a depth of 1 m. Active layer (approximately 35cm depth) and permafrost (approximately 85 cm depth) samples were cut from the frozen cores using a band saw in a −20 °C cold room. Approximately 10 g of frozen soil were incubated at 5 °C for one week under a helium headspace. CH_4 and CO_2 concentrations were measured on days 2 and 7 of incubation. In separate experiments, CH_4 oxidation was measured in permafrost samples incubated with 2-bromoethane sulphonic acid and CH_4 (1500 p.p.m.). CH_4 was subsequently measured in the headspace every 24 h for 7 days.

Shotgun metagenomic sequencing was conducted on DNA extracted from two replicate cores (active layer and permafrost) while frozen and after 2 and 7 days of incubation. Because the DNA yield was too low to prepare standard sequencing libraries, we used emPCR[13,19], for amplification before sequencing. This method escapes biases common in other amplification methods. The DNA was sequenced using the Illumina GAII platform with 2 × 113-base-pair (bp) cycles. Metagenomic reads were annotated by comparison with the KEGG genes database[20] using BLASTX[21] ($E \leq 1 \times 10^{-5}$) and with the Greengenes[22] and SILVA[23] databases using BLASTN[21] ($E \leq 1 \times 10^{-20}$). For assembly, raw reads from all samples were combined. After quality filtering and de-replication, *k*-mers (32-mers) with a depth less than two were removed from the data set[14]. The filtered reads were then assembled using Velvet[24]. Approximately ten assemblies were generated, and the one generating the longest contigs was chosen for further analysis. Contigs were annotated through the Integrated Microbial Genomes & Metagenomics (IMG/M) system[25]. To group contigs into a draft genome, large contigs (greater than 15 kb in length) were clustered based on tetranucleotide frequency and coverage. The largest bin was grown by the recruitment of smaller contigs to the large cluster[14].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 April; accepted 20 September 2011.
Published online 6 November 2011.

1. Schuur, E. A. G. *et al.* Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *Bioscience* **58,** 701–714 (2008).
2. Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23,** GB2023 (2009).
3. Zimov, S. A., Schuur, E. A. & Chapin, F. S. Climate change. Permafrost and the global carbon budget. *Science* **312,** 1612–1613 (2006).
4. Osterkamp, T. Characteristics of the recent warming of permafrost in Alaska. *J. Geophys. Res.* **112,** F02S02 (2007).
5. Prater, J. L., Chanton, J. P. & Whiting, G. J. Variation in methane production pathways associated with permafrost decomposition in collapse scar bogs of Alberta, Canada. *Glob. Biogeochem. Cycles* **21,** GB4004 (2007).
6. Chu, H. *et al.* Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ. Microbiol.* **12,** 2998–3006 (2010).
7. Hansen, A. A. *et al.* Viability, diversity and composition of the bacterial community in a high Arctic permafrost soil from Spitsbergen, Northern Norway. *Environ. Microbiol.* **9,** 2870–2884 (2007).
8. Steven, B., Pollard, W. H., Greer, C. W. & Whyte, L. G. Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high Arctic. *Environ. Microbiol.* **10,** 3388–3403 (2008).
9. Yergeau, E., Hogues, H., Whyte, L. G. & Greer, C. W. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. *ISME J.* **4,** 1206–1214 (2010).
10. Waldrop, M. P. *et al.* Molecular investigations into a globally important carbon pool: permafrost-protected carbon in Alaskan soils. *Glob. Change Biol.* **16,** 2543–2544 (2010).
11. Rivkina, R. *et al.* Biogeochemistry of methane and methanogenic archaea in permafrost. *FEMS Microbiol. Ecol.* **61,** 1–15 (2007).
12. Trotsenko, Y. A. & Khmelenian, V. N. Aerobic methanotrophic bacteria of cold ecosystems. *FEMS Microbiol. Ecol.* **53,** 15–26 (2005).
13. Blow, M. *et al.* Identification of ancient remains through genomic sequencing. *Genome Res.* **18,** 1347–1353 (2008).
14. Hess, M. *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331,** 463–467 (2011).
15. Sakai, S. *et al. Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order Methanocellales ord. nov. *Int. J. Syst. Evol. Microbiol.* **58,** 929–936 (2008).
16. Murray, P. A. & Zinder, S. H. Nitrogen-fixation by a methanogenic archaebacterium. *Nature* **312,** 284–286 (1984).
17. Johnson, S. S. *et al.* Ancient bacteria show evidence of DNA repair. *Proc. Natl Acad. Sci. USA* **104,** 14401–14405 (2007).
18. Ferry, J. G. How to make a living by exhaling methane. *Annu. Rev. Microbiol.* **64,** 453–473 (2010).
19. Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nature Methods* **3,** 545–550 (2006).
20. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
22. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).
23. Pruesse, E. C. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
24. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
25. Markowitz, V. M. *et al.* IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36,** D534–D538 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The work conducted by the Lawrence Berkeley National Laboratory Earth Sciences Division (Laboratory Directed Research Development) and the Joint Genome Institute was supported in part by the Office of Science of the US Department of Energy under Contract no. DE-AC02-05CH11231. This study was also supported by the Venture Capital and Yukon River Basin project of the United States Geological Survey. We acknowledge the technical support by the Joint Genome Institute production team. We thank A. Sczyrba, R. Egan and S. Canon for discussions and advice.

Author Contributions J.K.J., M.P.W. and K.M.D. conceived the incubation experiments. M.P.W. collected the samples. M.P.W. and S.J.B. conducted the incubation experiments. R.M., K.L.C. and K.M.D. performed the DNA extractions. R.M. created the shotgun sequencing libraries and conducted bioinformatics analyses. R.M. and K.M.D. performed statistical analyses. R.M. and M.M.D. performed qPCR experiments. R.M. and J.K.J. wrote the paper. All authors discussed the results and commented on the manuscript. E.M.R. M.P.W. and J.K.J. obtained funding for the study.

Author Information All annotated assembled sequences were incorporated into the IMG/M system with the Taxon Object ID 2067725009. Raw Illumina and 454 pyrotag sequence reads and a list containing the subset of contigs belonging to the draft methanogen genome are available at https://www.jgi.doe.gov/downloads/Permafrost_metagenome. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.K.J. (jrjansson@lbl.gov).

METHODS

Soil sampling. We obtained soil cores from a site in Interior Alaska that represents a C-rich form of permafrost. This site is a lowland soil near Hess Creek, north of Fairbanks, Alaska, and south of the Yukon River off the Dawson Highway (65° 40′ 12.84′′ N, 149° 04′ 36.24′′ W). The Hess Creek soil contains large quantities of organic matter in both the active layer and permafrost depths[10]. The ambient temperature of the soil was approximately −2 °C. Both sites have an overstory of black spruce (*Picea mariana*).

Soil cores were collected in March 2007 to a depth of 1 m while the soils were still frozen to the surface. The permafrost table began at 63 cm depth. Observed variability of active layer thickness is generally low in Central Alaska. Taking into account that the present-day conditions are among the warmest during the past 2000–3000 years, it is unlikely that thawing reached the sampled permafrost depths (V. Romanovsky, personal communication). We collected three replicate cores within an approximately 100 m^2 area. In the field, cores were scraped free of surface organic matter contaminants that froze to the core during the drilling process, wrapped in aluminium foil, placed into 4-inch diameter PVC tubes and capped. Cores were taken to the Cold Regions Research and Engineering Laboratory at Fort Wainwright Army Base in Fairbanks, Alaska. In a large −20 °C cold room, cores were again scraped of surface organic matter with sterile blades and cut into segments using a band saw. We selected sub-samples for sequencing according to depth: active layer (Core 1: 25–35 cm; core 2: 20–30 cm; core 3: 20–30 cm) and permafrost horizon (core 1: 75–85 cm; core 2: 75–100 cm; core 3: 85 cm–95 cm). Metadata collection and physical and chemical data for these samples are described elsewhere[10]. Briefly, these data showed that nutrients and labile C fractions were available to support microbial growth.

Incubation experiments. Frozen soil cores were cut using a band saw to excise experimental soil aliquots. Ice lens features and historical measurements at the sites indicate that the permafrost surface was at approximately 50–60 cm. Approximately 10 g of intact frozen soil were added to autoclaved glass jars (237.5 ml). The jars were sealed with septa fitted tops and were flushed with high purity He at 15 p.s.i. for 45 s and were then incubated at 5 °C in the dark. Ten millilitres of headspace were taken from the jars immediately after the soil was added and after 2 and 7 days incubation during which the soil thawed. The gas samples were injected into helium-cleared gas vials for methane measurements by gas chromatography–flame ionization detector. CO_2 concentrations in the headspace gas were measured at time zero, immediately after being placed in the cold room, and on days 2 and 7 using 2-ml syringe injections of headspace gas into an infrared gas analyser. All measurements were performed in triplicate.

For the CH_4 oxidation experiments, 5 g of soil was added to autoclaved glass jars (120 ml). Bromoethanesulphonate (BES) (190 mM) was added to a subset of the samples to a final concentration of 10 mM. An equivalent volume of sterile H_2O was added to non-BES incubations. The jars were sealed with septa fitted tops and flushed with helium at for 45 s. Methane was added to all samples with a final average concentration of 1500 p.p.m. and they were incubated at 5 °C in the dark. Headspace CH_4 was measured every 24 h for 7 days by gas chromatography–flame ionization detector. All treatments were incubated in triplicate.

DNA extraction. Three DNA extractions per sample were conducted according to a modified bead-beading protocol[26–28]. Approximately 0.5 g of soil was added to 2.0-ml tubes containing 1.4-mm ceramic spheres, 0.1-mm silica spheres, and one 4-mm glass bead (MP Biomedicals). Hexadecyltrimethylammonium bromide (CTAB) extraction buffer containing 10% CTAB in 1 M NaCl, 0.1 M ammonium sulphate and 0.5 ml phenol:chloroform:isoamylalcohol (25:24:1) was added and shaken in a FastPrep Instrument (MP Biomedicals) at 5.5 m s^{-1} for 30 s. After bead beating, the samples were extracted with chloroform and precipitated in a PEG 6000/1.6 M NaCl solution. Pellets were washed with 70% ethanol and re-suspended in molecular biology grade water. The three extractions were combined at this step and purified using an AllPrep DNA/RNA kit (Qiagen). Extracts were quantified using Quant-iT dsDNA HS assay kit (Invitrogen) according to the manufacturer's manual.

16S rRNA gene sequencing and analysis. Each sample was amplified with the primer pair 926f/1392r as described in Kunin *et al.*[29]. The reverse primer included a 5-bp barcode for sample multiplexing during sequencing. After PCR, samples were purified using a MiniElute PCR purfication kit (Qiagen) and quantified using the Agilent Labchip System. Samples were then pooled at equal concentrations. Sequencing of the amplicons was performed using Roche 454 GS FLX according to the manufacturer's instructions.

Sequences were analysed using the software tool PyroTagger[30]. Sequences that were less than 220 bp in length and those with greater than 3% low quality bases (quality score less than 27) were removed. Unique sequences were clustered at 97% identity using the Uclust option[31]. Cluster representatives were classified using Greengenes[22]. The phylogeny of cluster representatives was inferred using an approximate maximum likelihood method designed for large alignments as implemented by the software FastTree[32]. The weighted UniFrac metric[33] was used to quantify differences in community composition. Statistical tests for differentially abundant operational taxonomic units were made using the Metastats methodology[34] with 1,000 permutations to compute non-parametric *P* values.

Emulsion PCR paired-end library generation. Starting material for library generation ranged from 1 to 10 ng depending on the total amount of DNA obtained from the extractions. The DNA was sheared to 300–500 bp in 100-µl round-bottom glass tubes on a Covaris-S instrument (Covaris). Ends were repaired using the End-It DNA Repair Kit (EPICENTRE Biotechnologies) according to the manufacturer's instructions. End-repaired DNA was purified using phenol:chloroform:isoamylalcohol (25:24:1), precipitated in ethanol and resuspended molecular biology grade water.

Illumina paired-end linkers were ligated to the end-repaired DNA by adding 1 µl PEG, 1 µl 10× ligation buffer, 1 µl T4 DNA ligase (5 U $µl^{-1}$) (Fermentas), 20 µM each of the Illumina paired-end adapters, and incubated at 22 °C for 1 h. DNA was purified using a MiniElute Reaction Cleanup kit (Qiagen).

Standard library creation protocols were not used owing to low DNA yield (only a few nanograms per sample). Therefore, Illumina paired-end libraries were made by linker-mediated emPCR amplification as described previously[13,19]. This method theoretically escapes bias common in other methods, such as multiple displacement amplification, by isolating DNA molecules in aqueous droplets[15]. Briefly, a PCR mix was created using the linker-ligated DNA, 10× PfU PCR buffer, BSA, dNTPs, primers, water and Pfu Turbo DNA polymerase (Stratagene). Fifty microlitres were reserved as a non-emulsion control and the rest was added dropwise to 400 µl of an oil–surfactant mixture. The resulting emulsion was subjected to 40 PCR cycles. The emulsion was broken by addition of diethyl ether and DNA was recovered using chloroform and subsequently by using a MiniElute PCR Purification kit. Products were run on a 2% low melt agarose gel and 250- to 500-bp-sized products were cut from the gel and purified using a MiniElute Gel Extraction Kit (Qiagen). Libraries were quantified on a Bioanalyser DNA 1000 chip (Agilent). Sequencing was performed according to the manufacturer's instructions (Illumina). The flow cell was sequenced using Illumina GAII technology generating 2 × 113-bp paired end reads with an average insert size of approximately 300 bp.

Annotation of metagenomic sequence reads. Metagenomic reads were annotated through comparison with the KEGG database[20] using BLASTX[21] at an E value cutoff of 1×10^{-5}. The relative abundance of each gene was determined by taking the number of hits to that gene in a given sample and dividing by the total number of hits to the KEGG gene databases. Small subunit ribosomal RNA gene sequences in the metagenome were identified through BLASTN[21] by comparison with the Greengenes database[22] (16S) and the Silva database[23] (18S) at an E value cutoff of 1×10^{-20}. Methane monooxygenase (*mmo*) genes were identified through additional comparisons to all known *mmo* genes in the NCBI nt and nr non-redundant databases (BLASTX, E value cutoff 1×10^{-5}). Each putative *mmo* read was checked against the entire KEGG genes database and was discarded if it matched another gene with a higher bit score.

Statistical analyses. Ordination based on the relative abundance of KEGG genes was performed using principal component analysis. Identification of individual genes differing in relative abundance between the frozen active and permafrost layer samples was conducted using a binomial test implemented in the R package[35] ShotgunFunctionalizerR[36]. Genes were considered enriched in a given layer if the relative abundance was significantly different between layers from both cores. *P* values were corrected for multiple tests using the Benjamini–Hochberg correction factor[37].

For comparisons between the pyrotag and metagenome data, phylogenetic information was extracted from the metagenomes using MEGAN[38]. Hits corresponding to specific taxa were retained if their bit scores were within 10% of the best bit score, and if the minimum score was 35 as suggested for short reads. Additionally, singleton hits were excluded from the analysis. For analysis of pyrotag data compared with the metagenomic data, a subset of the data was made of only the samples whose metagenome were also sequenced ($n = 12$). Based on this subset, any taxa (cluster) was excluded that was only represented once, as was the case for the metagenome data set. These analyses were performed in R[35]. Ordination of taxonomic community data was performed using the ordination method non-metric multi-dimensional scaling with a Bray–Curtis distance measure. Where appropriate, *P* values were corrected for multiple tests using the Benjamini–Hochberg correction factor. Non-parametric analysis of variance based on dissimilarities was performed using adonis function in the vegan package of R[39]. Hierarchical clustering was performed using the R package pvclust[40], with 10,000 bootstrap replications.

Response of C and N cycle genes to thaw. To determine the change in genes relevant to C and N after thaw, we selected a subset of KEGG genes based on C and N cycle pathways. Difference in abundance was determined by comparing the

relative abundance at the frozen and day 7 time points and calculating the $\log_2$-fold change.

Metagenome assembly and annotation. All sequence data were combined, screened and trimmed using a *k*-mer-based filtering approach. *k*-mers with a depth less than two were removed as described previously[14] because they are not expected to contribute to the assembly and because the reduction reduces memory requirements for subsequent assembly steps. Filtered sequences were assembled by Velvet[24] from all reads using a *k*-mer length of 51, insertion length of 300 bp and an expected coverage of 100. Coverage was calculated by mapping all reads back to the contigs using BWA at default parameters[41]. The assembly used approximately 5% of de-replicated reads. All contigs were submitted to the US Department of Energy Joint Genome Institute IMG/M system (http://img.jgi.doe.gov/cgi-bin/m/main.cgi)[25] for gene calling and functional annotation.

BLASTN searches of all contigs were performed against the NCBI-nt non-redundant database. Assignment to phylogenetic groups at the level of class by was done using MEGAN[38] at default parameters. By this measure, approximately 80% of the large contigs (greater than 10 kb in length) originated from methanogenic taxa. Therefore, we focused our further assembly efforts on methanogens. Larger contigs were clustered by genome based on tetranucleotide frequencies and coverage, two properties expected to be present in contigs derived from the same genome[14]. Briefly, hierarchical agglomerative clustering using Euclidean distance (maximum distance: 0.13) was performed on all contigs greater than 15 kb in length. The largest cluster was manually inspected and contigs with a coverage that differed from the mean by more than one standard deviation were removed. The largest bin was grown by recruiting the remaining contigs (less than 15 kb) where the tetranucleotide frequency distance was less than 0.18. Coverage cutoff was set at one standard deviation from the mean.

For the single-copy gene analysis, we selected 19 protein-encoding genes (*rpl1p*, *rpl2p*, *rpl3p*, *rpl4lp*, *rpl5p*, *rpl6p*, *rpl11p*, *rpl13*, *rpl14p*, *rpl19e*, *rps2p*, *rps5p*, *rps9p*, *rps10p*, *rps11p*, *rps13p*, *dnaG*, *pyrG*, *nusA*) that were identified previously by Wu and Eisen[42] to be universally distributed in bacteria, exist as single-copy genes and are recalcitrant to lateral gene transfer. Because the study of Wu and Eisen focused on bacteria, we further confirmed that the selected genes exist unambiguously as single-copy genes in all methanogens with a sequenced genome. The occurrence of each gene in the draft genome was determined from the IMG/M functional annotations.

For phylogenetic analysis, we selected a subset of the 19 single-copy genes (*rpl1p*, *rpl2p*, *rpl3p*,*rpl5p*, *rpl6p*, *rpl19e*, *rps11p*). These genes were extracted from the draft genome and downloaded from 15 methanogen genomes (NCBI accession numbers NC_009712, NC_003551, NC_000916, NC_013790, NC_007681, NC_014658, NC_000909, NC_014002, NC_007955, NC_009464, NC_014570, NC_009051, NC_008942, NC_007796), selected to represent the diversity of known methanogens. DNA sequences were concatenated and then aligned using MUSCLE[31]. Low confidence parts of the alignments were trimmed using gblocks at default settings[43]. A phylogenetic tree comparing the draft genome with other methanogen genomes was inferred using FastTree[32].

For further validation of the draft genome, we selected a related reference genome determined by phylogentic analysis of the single-copy genes. The draft genome was compared with its sequenced relative (NC_009464) using BLASTN with an E value cutoff of 1×10^{-5} and the penalty for a nucleotide mismatch set at -1. A given base in the reference genome was considered covered if it was part of the alignment. The number of times each base in the reference genome was covered by draft genome was counted. To determine the number of times any given base is expected to be covered when comparing two finished genomes, the same analysis was conducted comparing NC_009464 with two genomes (NC_007955 and NC_014002) that were found to be approximately as distant to the reference genome as the draft genome.

Quantitative PCR. We used qPCR to quantify the abundances of *mcrA*, *nifH*, *narG*, *pmoA* functional genes, and the 16S rRNA genes of type I and type II methanotrophs. Experiments were performed on unamplified DNA samples in 25 μl volumes. The PCR conditions and primers are given in Supplementary Table 7. We used the QuantiTect SYBR Green PCR kit (Qiagen). All PCRs were run on an iCycler machine (Bio-Rad Laboratories). Reactions consisted of 2 μl of template DNA (diluted to 2 ng μl^{-1}) and 1 μl each of the forward and reverse primers (10 μM). Standards were created by amplifying the gene fragment of interest from a reference genome (in the case of *mcrA*, *nifH*, *narG* and the 16S rRNA gene sequences) or by synthesis of the positive control (in the case of *pmoA*) as described previously[44]. PCR products resulting from the amplification were cloned using a Zero Blunt TOPO TA cloning kit and One Shot TOP 10 chemically competent cells (Invitrogen).

26. DeAngelis, K. M. *et al.* Selective progressive response of soil microbial community to wild oat roots. *ISME J.* **3,** 168–178 (2009).
27. Brodie, E., Edwards, S. & Clipson, N. Bacterial community dynamics across a floristic gradient in a temperate upland grassland ecosystem. *Microb. Ecol.* **44,** 260–270 (2002).
28. Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G. & Bailey, M. J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **66,** 5488–5491 (2000).
29. Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12,** 118–123 (2010).
30. Kunin, V. & Hugenholtz, P. PyroTagger: a fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. *Open J.* **1,** 1 (2010).
31. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
32. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26,** 1641–1650 (2009).
33. Hamady, M., Lozupone, C. & Knight, R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* **4,** 17–27 (2010).
34. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5,** e1000352 (2009).
35. R Development Core Team. R: A Language and Environment for Statistical Computing ⟨http://www.R-project.org⟩ (2010).
36. Kristiansson, E., Hugenholtz, P. & Dalevi, D. ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* **25,** 2737–2738 (2009).
37. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J.R. Stat. Soc. B* **57,** 289–300 (1995).
38. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17,** 377–386 (2007).
39. Oksanen, J. *et al.* Vegan: Community Ecology Package. R package version 1.15-4. ⟨http://CRAN.R-project.org/package=vegan⟩ (2011).
40. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22,** 1540–1542 (2006).
41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
42. Wu, M. & Eisen, J. A. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* **9,** R151 (2008).
43. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignment. *Syst. Biol.* **56,** 564–577 (2007).
44. David, M. M., Sapkota, A. R., Simonet, P. & Vogel, T. M. A novel and rapid method for synthesizing positive controls and standards for quantitative PCR. *J. Microbiol. Methods* **73,** 73–77 (2008).