# Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming

Kai Xue[1,2,3†], Mengting M. Yuan[2,3†], Zhou J. Shi[2,3], Yujia Qin[2,3], Ye Deng[2,3,4], Lei Cheng[2,3,5], Liyou Wu[2,3] , Zhili He[2,3], Joy D. Van Nostrand[2,3], Rosvel Bracho[6], Susan Natali[7], Edward. A. G. Schuur[6,8], Chengwei Luo[9], Konstantinos T. Konstantinidis[9], Qiong Wang[10], James R. Cole[10], James M. Tiedje[10], Yiqi Luo[3] and Jizhong Zhou[1,2,3,11*]

[1]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China. [2] Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma 73019, USA. [3] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, USA. [4] Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 10085, China. [5] College of Life Sciences, Zhejiang University, Hangzhou 310058, China. [6] Department of Biology, University of Florida, Gainesville, Florida 32611, USA. [7] Woods Hole Research Center, Falmouth, Massachusetts 02540, USA. [8] Center for Ecosystem Sciences and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86001, USA. [9] School of Civil and Environmental Engineering, and School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. [10] Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824, USA. [11] Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94270, USA. [†]These authors contributed equally to this work. *e-mail: jzhou@ou.edu

## Abstract

Microbial decomposition of soil carbon in high-latitude tundra underlain with permafrost is one of the most important, but poorly understood, potential positive feedbacks of greenhouse gas emissions from terrestrial ecosystems into the atmosphere in a warmer world[1,2,3,4]. Using integrated metagenomic technologies, we showed that the microbial functional community structure in the active layer of tundra soil was significantly altered after only 1.5 years of warming, a rapid response demonstrating the high sensitivity of this ecosystem to warming. The abundances of microbial functional genes involved in both aerobic and anaerobic carbon decomposition were also markedly increased by this short-term warming. Consistent with this, ecosystem respiration ($R_{eco}$) increased up to 38%. In addition, warming enhanced genes involved in nutrient cycling, which very likely contributed to an observed increase (30%) in gross primary productivity (GPP). However, the GPP increase did not offset the extra $R_{eco}$, resulting in significantly more net carbon loss in warmed plots compared with control plots. Altogether, our results demonstrate the vulnerability of active-layer soil carbon in this permafrost-based tundra ecosystem to climate warming and the importance of microbial communities in mediating such vulnerability.

## Introduction

Permafrost, defined as 'subsurface earth materials remaining below 0 °C for two consecutive years'[1], is a unique characteristic of polar regions and high mountains. In permafrost-underlain high-latitude tundra, plant-derived carbon has accumulated over hundreds to thousands of years because low temperatures and saturated soils reduce microbial decomposition of soil organic C (refs 5,6). As a result, nearly 50% of the global soil organic C is stored in Northern Hemisphere permafrost and the active-layer soils above, although they cover only 16% of the global terrestrial area[7]. High-latitude tundra has long been recognized as being highly responsive to climate change[8]. Recent accelerated warming in the northern high-latitude region[9] has resulted in rapid permafrost degradation, and studies suggest that permafrost could decline by 30–70% by the end of the twenty-first century[10,11]. During permafrost degradation, frozen soil becomes biologically active, with microbial decomposition resulting in massive ecosystem C loss, which is likely to dominate the overall net C exchange in permafrost regions[1]. Although plant responses to climate warming in the active layer of the tundra soil have been intensively studied[12,13,14], microbial responses have not been examined until very recently[4,15,16,17,18].

Although various observational studies have documented the responses of tundra ecosystems to natural warming[19], and some incubation studies revealed microbial community changes on permafrost thaw in laboratory settings[16,18], very few studies examined microbial responses to climate warming in tundra ecosystems in the field. As field experimental warming can directly examine the impacts of temperature increases on the microbial community *in situ*[13], an ecosystem warming experiment, Carbon in Permafrost Experimental Heating Research (CiPEHR), was established in September, 2008, in Interior Alaska. The experiment is located in typical moist acidic tussock tundra[20], a dominant tundra type, on permafrost that is close to the freezing point and thus especially vulnerable to thaw in a warming climate[21]. In this experiment, snow fences (that is, increased snow pack for insulation) were used in the soil warming treatment to increase soil temperature, coupled with early spring snow removal to control snow-water equivalents in both warmed and control plots. Soil warming and control treatments were arranged in six replicates, providing sufficient statistical power. This is the first warming experiment to degrade surface permafrost without delaying spring snow melt[14]. To understand how vulnerable the active layer of the tundra soil is to climate warming, a total of 12 subsurface soil samples from a representative depth of 15–25 cm were collected from both warmed and control plots after short-term (1.5 years) warming for geochemical and microbial analyses. These samples represented active-layer soil that freezes in winter and thaws in the growing season, and were within the organic horizon along the depth profile. As soil microbial community structure is tightly linked to changes in the aboveground plant community and soil environmental conditions[22], we predicted that short-term warming

would result in selective microbial growth, which would be seen as a shift in the active-layer microbial community structure and accompanying gene content, especially in those populations and traits important to both aerobic and anaerobic C decomposition and nutrient cycling. Consequently, soil C in this tundra ecosystem would be highly vulnerable to climate warming.

Short-term soil warming altered several environmental attributes (for example, plant, soil microclimate and soil properties) of the tundra[12,14,23]. First, soil temperature (5–40 cm) increased by 2.3 °C (from −6.2 °C to −4.0 °C) in response to warming in wintertime and by 0.6 °C (from 3.8 °C to 4.4 °C) during the growing season in 2010 (Fig. 1a), which led to a substantial surface permafrost thaw as indicated by an increased thaw depth (8.8%, $p < 0.001$; Fig. 1c). Similarly, soil moisture increased in response to warming (over 10%, $p = 0.03$; Fig. 1b). Second, GPP increased (30.3%, $p = 0.02$; Fig. 1f), mainly owing to enhanced growth of graminoids (57.5% increase in biomass, $p = 0.05$). Warming also extended the growing season length through earlier bud break and delayed senescence[12]. In addition, the percentage of mostly cellulose fraction of the labile C pool[24] in total soil organic C was higher (36.1%, $p = 0.06$) in warmed than control soils (Fig. 1e). The C amount of this pool under warming tended to increase as well, but not statistically significantly (Supplementary Fig. 5). Together, these results indicated that environmental attributes of the tundra soil were altered rapidly by short-term warming.

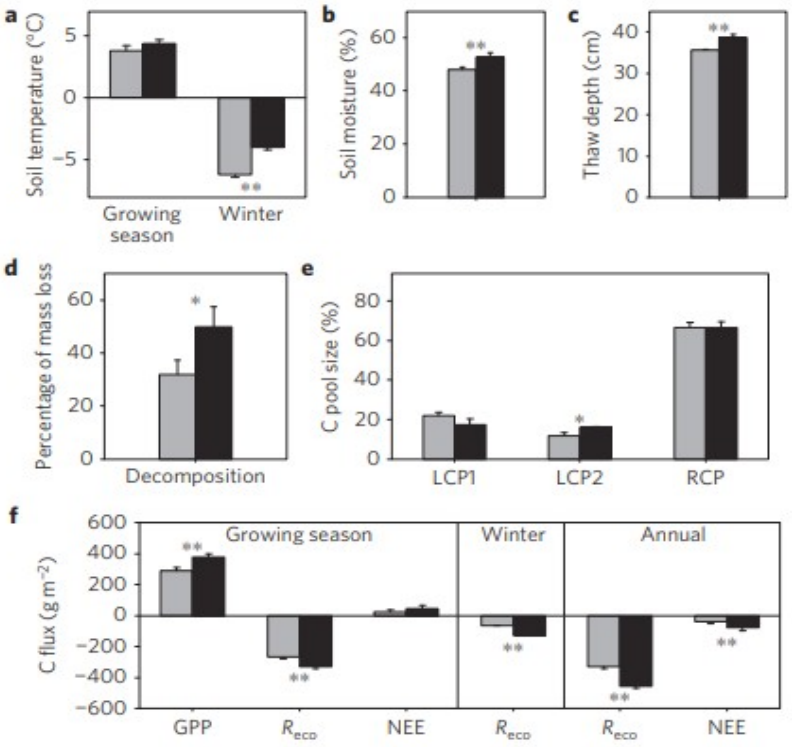


**Figure 1 | Warming effects on soil variables and ecosystem C fluxes.** Grey bars represent control plots and black bars represent warmed treatment plots. **a**, Soil temperature in both growing season (May to September 2010) and wintertime (December 2009 to March 2010) averaged across 5, 10, 20 and 40 cm. **b**, Soil moisture. **c**, Maximum thaw depth. **d**, Standard cellulose filter paper decomposition rate (mass loss) in the field. **e**, Proportion of soil C pools in total organic C, including labile C pool 1 (LCP1, mainly polysaccharides) and 2 (LCP2, mostly cellulose), and recalcitrant C pool (RCP). **f**, Growing season (May to September 2010), wintertime (October 2009 to April 2010) and annual ecosystem C fluxes, which were estimated on the basis of the C amount from $CO_2$ emissions. GPP, gross primary productivity; $R_{eco}$, ecosystem respiration; NEE, net ecosystem C exchange. Positive values indicate C sink, and negative values represent C source. Error bars represent standard error of the mean. The differences between warmed and control plots were tested using two-tailed *t* tests, indicated by ** when $p < 0.05$, or * when $p < 0.10$. Panels **a**–**d**,**f** were reanalysed from previously published data[12,23,31].

The observed alterations in the soil microclimate (temperature, moisture, thawing depth), soil C, and GPP in response to warming would be expected to cause significant changes in the microbial communities in the active layer of tundra soil. Consistent with this expectation, the microbial community functional gene structure was markedly different between warmed and control plots as revealed by the detrended correspondence analysis of the GeoChip data (Supplementary Fig. 1a), indicating increases in certain genes and possibly the organisms that host these. Three different non-parametric multivariate statistical tests (ANOSIM, Adonis and MRPP) showed that the functional community structure differed substantially between warmed and

control plots (Table 1). However, significant differences in the overall community structure were not detected with 16S ribosomal RNA gene-based amplicon and shotgun metagenomic sequencing approaches (Supplementary Fig. 1b, c). This is most likely due to the high heterogeneity of soil environments, low taxonomic resolution of the experimental approaches, and/or high noise associated with random sampling25. Canonical correspondence analysis revealed that soil temperature, moisture and plant GPP were the main significant variables related to the microbial community functional structure ($F = 1.68$,$p = 0.005$; Supplementary Fig. 2a). This is also consistent with our central hypothesis that warming-induced changes in plant productivity and soil microclimate significantly alter the soil microbial community structure. In addition, soil community DNAs were shotgun sequenced and a total of 3.24 billion raw sequences were obtained for these samples (Supplementary Table 2). Although the overall metagenome structures were not separable into warmed versus control groups (Supplementary Fig. 1c, Supplementary Fig. 3 and Supplementary Table 1), a small portion (7.4%) of total subsystems, genes associated with microbial physiological attributes and ecosystem processes, were significantly different between warmed and control plots ($p < 0.05$; Fig. 2d). In particular, warmed plots were enriched in genes associated with pathways related to labile C utilization (Supplementary Table 3). Our above results indicated that the microbial communities in the active layer of tundra soil were responsive to warming.

**Table 1 | Significance tests on the effects of warming on the microbial community functional structure detected by GeoChip hybridization.**

| | MRPP | | ANOSIM | | Adonis | |
|---|---|---|---|---|---|---|
| | $\delta$ | $p$ | $R$ | $p$ | $F$ | $p$ |
| Euclidean | 69.02 | **0.04** | 0.38 | **0.02** | 1.69 | **0.04** |
| Horn | 0.06 | **0.04** | 0.36 | **0.02** | 1.20 | 0.23 |
| Bray | 0.09 | **0.03** | 0.30 | **0.02** | 1.28 | 0.21 |

Three different permutation tests were performed, including the multiple response permutation procedure (MRPP), analysis of similarity (ANOSIM) and permutational multivariate analysis of variance (Adonis), calculated with Euclidean, Horn or Bray distance. Bold values indicate $p < 0.05$.

Soil warming also significantly impacted a number of microbial functional and phylogenetic groups important for C decomposition. First, more than half (54.5%) of the detected C decomposition genes were increased by warming according to GeoChip signal intensities ($p < 0.05$; Fig. 2a), including those involved in degrading starch (for example, *amyA* encoding $\alpha$-amylase), hemicellulose (for example, *ara* encoding arabinofuranosidase), cellulose (for

example, cellobiase), chitin (for example, endochitinase), aromatics (for example, *vdh* encoding vanillin dehydrogenase) and lignin (for example, glyoxal oxidase, phenol oxidase). Also, the total fungal functional gene intensity detected by GeoChip was more abundant in warmed plots than control plots (4.7%, $p < 0.001$; Supplementary Fig. 4a). Increases of the genes involved in recalcitrant C decomposition (Fig. 2a) suggest the possible degradation of old recalcitrant C and thus a potential positive feedback to climate warming. In addition, shotgun metagenome sequence data revealed that a substantial portion (19.5%, 8 of 41) of C degradation pathways was increased by warming ($p < 0.05$; Fig. 2d, Supplementary Table 3), including those for cellulose, mannose metabolism, carbohydrate hydrolases, fructooligosaccharides and raffinose utilization, lactose and galactose uptake and utilization, L-fructose utilization, xylose utilization, chitin utilization and *N*-acetylglucosamine utilization. More specifically, many individual genes involved in labile C degradation (for example, starch, hemicellulose and cellulose), which were identified from metagenome sequences using GeoChip genes as queries, were increased (95% confidence interval; Fig. 2e). Overall, as these functional genes directly participate in aerobic C degradation, their higher abundance could enhance C decomposition and hence contribute to positive climate feedback.

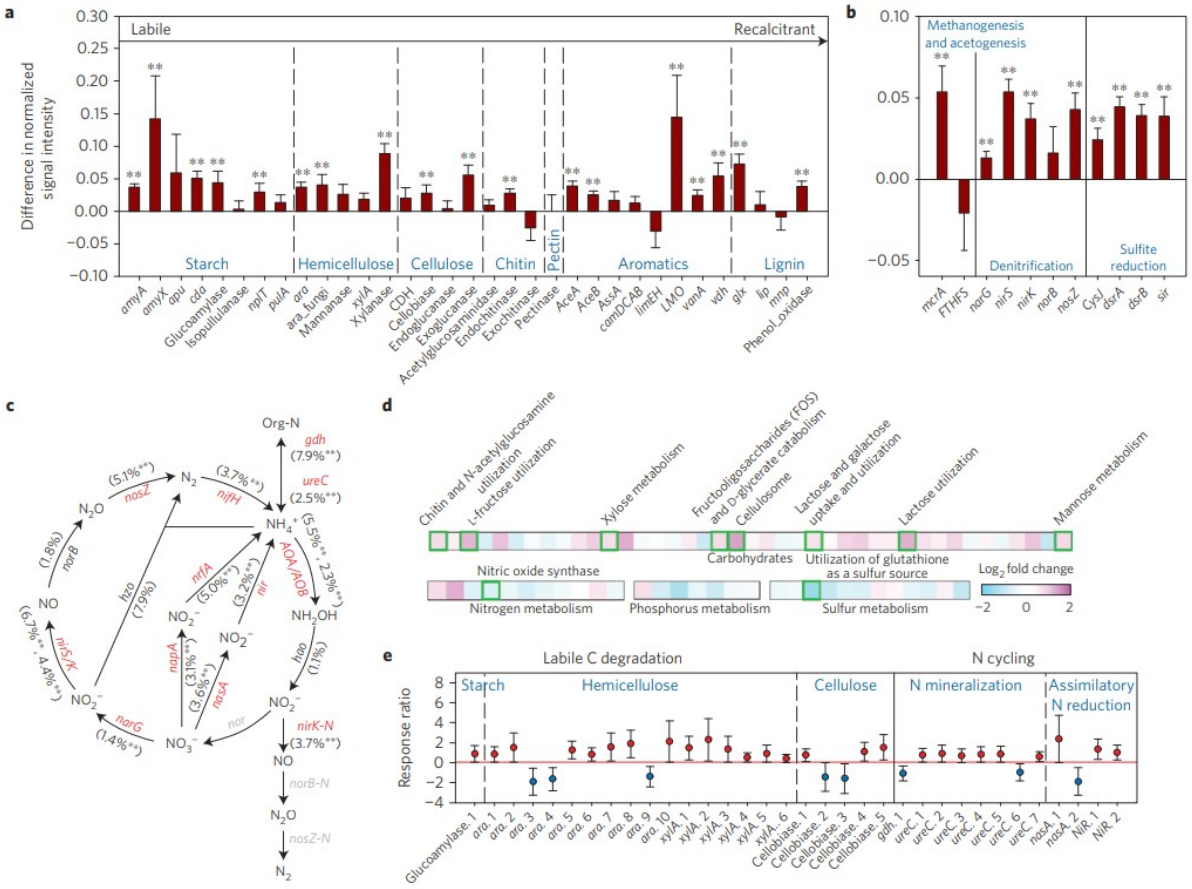


**Figure 2 | Warming effects on functional genes involved in biogeochemical cycling processes. a**, C degradation from GeoChip data. The targeted substrates were arranged in order from labile to recalcitrant C. GeoChip data are presented as the signal difference between warmed and control plots (warmed–control). Error bars represent standard error. Significance is indicated by ** when $p < 0.05$. **b**, Anaerobic processes from GeoChip data. **c**, N processes from GeoChip data. The percentage change in N gene abundance in response to warming is indicated in parenthesis. Genes where change in abundance was significant ($p < 0.05$) are labelled in red. Grey-coloured genes were not targeted by the version of GeoChip used here, not detected or not applicable. **d**, Abundance of subsystems involved in C, N, phosphorus and sulfur cycling from metagenomic shotgun sequence data. Changes in subsystems are indicated as fold change ($\log_2$ (warmed/control)) in abundance. Significant differences between warmed and control plots are highlighted with green squares. **e**, Response ratios showing significant changes in abundance of gene clusters involved in C and N cycling from the metagenomic shotgun sequence data. These gene clusters were identified by searching the shotgun sequence data sets using GeoChip genes as queries. Each cluster on the *x* axis represents a group of sequences among which the similarities are ≤95%. The GenBank GI numbers of the representative sequences for the gene clusters are listed in Supplementary Table 8. Error bars indicate 95% confidence intervals of abundance differences between warmed and control groups. The full names of the genes in this figure are listed in Supplementary Table 7.

The potential for accelerated C decomposition was supported by several independent pieces of evidence. First, the cellulose decomposition rate measured by adding external cellulose substrate was higher under warming (Fig. 1d). Also, both winter $R_{eco}$, derived almost exclusively from heterotrophic soil respiration, and growing season $R_{eco}$, including both autotrophic (from plants) and heterotrophic soil respiration, increased with warming (100% and 24%, $p < 0.05$; Fig. 1f; ref. 23). In addition, strong correlations were observed between both growing season and wintertime $R_{eco}$ and the functional gene groups involved in degrading almost all C compounds targeted by GeoChip, including starch, hemicellulose, cellulose, chitin, aromatics and lignin (Supplementary Table 5), suggesting that changes in abundance of these genes could be important in mediating $R_{eco}$ (ref. 26).

Permafrost thawing induced by warming often increases soil water content and creates a mosaic of flooded areas interspersed within dry areas, which may potentially enhance anaerobic C decomposition[1,18]. As water is heterogeneously distributed spatially and temporally, oxygen would also be expected to be unevenly distributed in the soil[14]. GeoChip hybridization-based analysis revealed that genes involved in several important anaerobic respiration processes, such as denitrification, methanogenesis and sulfate reduction, were increased by warming ($p < 0.05$; Fig. 2b). Shotgun metagenome sequence analysis also revealed that the pathway for denitrification (marginally) increased in response to warming ($p = 0.08$; Supplementary Table 6). These results were consistent with laboratory incubation studies, where methanogenic pathways were increased within several days after permafrost thaw[16,18]. Although some upland permafrost areas were observed to be $CH_4$ sinks[26], for our studied site, significantly increased $CH_4$ emission rates after several years of warming have been reported[27]. As $CH_4$ and $N_2O$ have 28 and 265 times the warming potential of $CO_2$ per mole[28], respectively, anaerobic microbial responses are most likely of considerable importance.

Estimates show that methane emission in northern wetlands, including tundra, accounts for 25% of the global methane release from natural sources[29]. GeoChip analysis revealed that warming increased the gene encoding methyl coenzyme M reductase A (*mcrA*), a key enzyme in methanogenesis ($p < 0.01$; Supplementary Fig. 4b). Although 16S rRNA sequence analysis detected two methanogens (*Methanobacteria* and *Methanomicrobia*; Supplementary Fig. 4d), the relative abundance between warmed and control plots was not significantly different for either methanogen. Warming, however, resulted in a greater abundance of *pmoA*, a gene encoding particulate methane monooxygenase subunit A ($p = 0.01$; Supplementary Fig. 4b), suggesting that more of the methane produced could be oxidized in the aerobic upper soil horizon at the warmed plots. Similar findings were reported in recent studies on both incubated permafrost soils[16] and active-layer samples *in situ*[17,26].

Warming also increased genes involved in N cycling (Fig. 2c, e), microbial phosphorus utilization and sulfur metabolism (Supplementary Fig. 4c). Most (82.4%) of the GeoChip-detected functional genes involved in N cycling were increased in response to warming ($p < 0.05$; Fig. 2c), consistent with the previous finding that warming enhances nutrient cycling[3]. For example, the abundance of $N_2$-fixing bacteria was higher in response to warming ($p < 0.05$; Fig. 2c), and two bacterial classes (Opitutae and Deltaproteobacteria) detected by PCR amplification of the *nifH* gene had higher abundance in warmed samples ($p < 0.05$; Supplementary Fig. 4e). Also, the abundance of key genes (for example, *gdh* and *ureC*) in N mineralization was higher in warmed than control soil ($p < 0.05$ or 95% confidence interval; Fig. 2c, e). In addition, warming seemed to increase nitrification and denitrification processes, as indicated by increased *nirK* and *amoA* genes from GeoChip

data (Fig. 2c). The increase in *amoA* could potentially lead to higher nitrate concentrations, which is also supported by the greater abundance of genes for various reductive processes that use nitrate as an electron acceptor, such as *narG*, *nirS/nirK* and *nosZ* for denitrification, *napA* and *nrfA* for dissimilatory nitrate reduction to ammonium, and *nasA*, *nirA* and *nirB* for assimilatory nitrate reduction (Fig. 2c, e). Microbial phosphorus utilization genes (phytase and *ppx*) and eight of the eleven detected sulfur metabolic genes had higher abundance in warmed than in control plots (Supplementary Fig. 4c). Although the significant increase in abundance of the genes involved in nutrient-cycling processes observed in warmed plots may potentially enhance the rates of nutrient cycling, more in-depth studies are necessary to determine the rates and extent of stimulation of different nutrient-cycling processes.

The increased abundance of N cycling genes (particularly those involved in N mineralization, N fixation and nitrification) and other nutrient-cycling genes could increase nutrient (especially N) availability in soil, which is important for ecosystem C dynamics because N is a limiting factor for plant growth in most tundra ecosystems[12]. That warming enhanced plant N uptake is supported by the observation that from 2009 to 2010 plant foliar N mass increased in warmed plots (35%, $p < 0.01$), but remained unchanged in control plots[12]. The enhanced plant N uptake could in turn affect GPP, which increased in response to warming ($p < 0.05$; Fig. 1f). Moreover, almost all genes involved in N cycling (18 of 19) and most C degradation genes (22 of 33) showed significant correlations with GPP ($p < 0.05$; Supplementary Table 5). Increased $N_2$ fixation, mineralization and nitrification could counteract the potential higher N loss from soil due to increased plant N uptake, denitrification and nitrate leaching. As a net result, the soil N availability seemed not to be affected by warming[14].

In summary, our results highlight the importance of microbial-community-mediated feedbacks of the active layer to warming, as illustrated in a conceptual model (Fig. 3). In response to warming, deeper thaw depth increased the amount of C accessible for decomposition. Within the active layer, soil C is also more vulnerable to degradation through the following mechanisms: first, short-term soil warming altered the active-layer microbial community structure, demonstrating rapid responses by these communities; annual $R_{eco}$ released 127 g more C $m^{-2}$ from warmed plots compared with controls, resulting in 38.6% more C loss from soil. Also, warming increased the abundance of functional genes involved in anaerobic processes, which could lead to a greater positive feedback by releasing more $CO_2$, $CH_4$ and $N_2O$ (Fig. 3). In contrast, the potentially higher nutrient availability resulting from the increased abundance of nutrient-cycling genes would also stimulate plant growth (Fig. 3). In this study, increased GPP did not completely offset the C loss from the warming-induced $R_{eco}$ increase. The net C loss from warmed plots doubled in 2010 (Fig. 1f), and was estimated to increase more in an actual climate warming scenario[23]. However, it should be noted that the

experimental results reported in this study were derived from the active layer of the Alaskan tundra soil. To generalize whether the results observed are applicable to permafrost requires further analyses with actual permafrost.

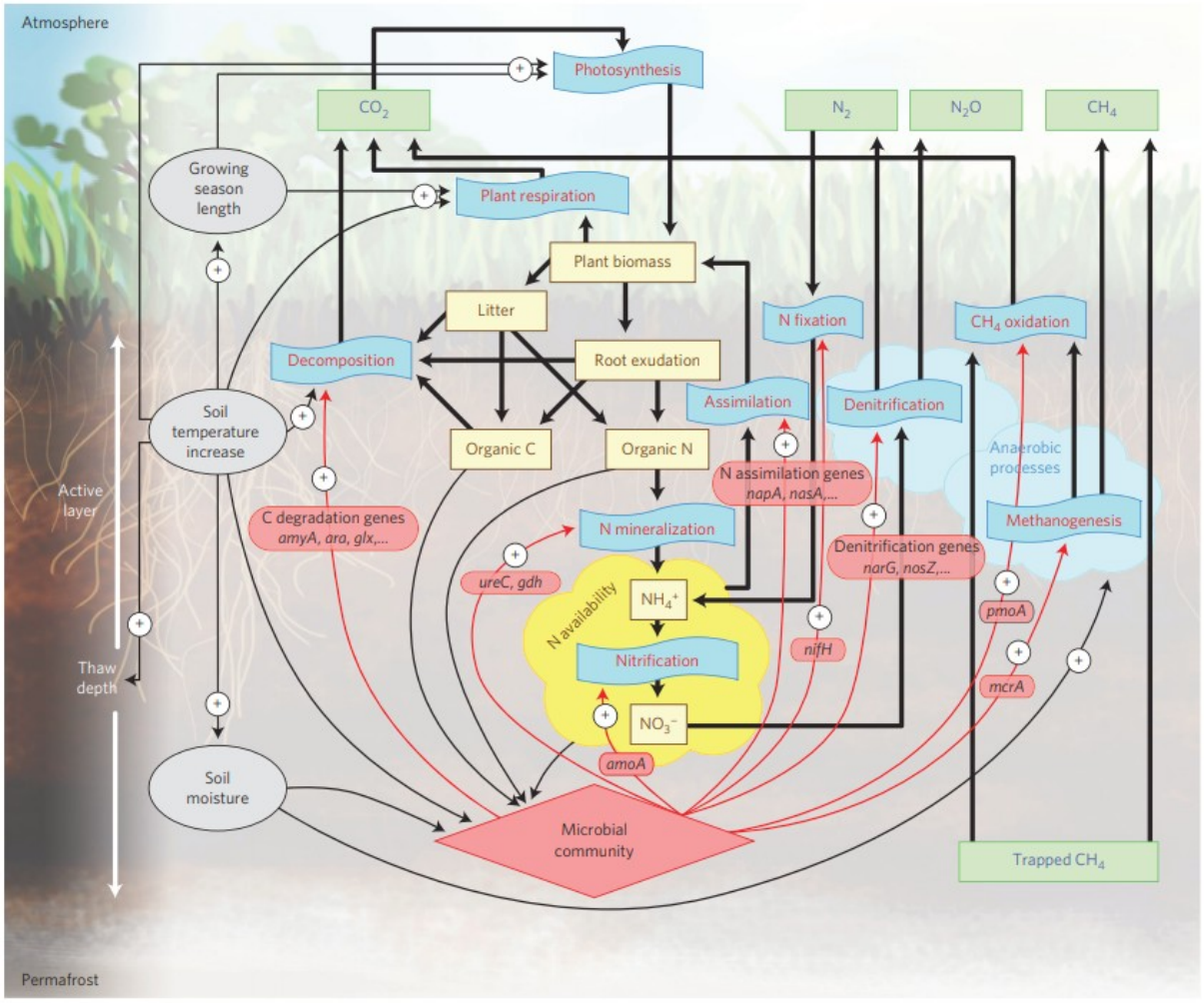


**Figure 3 | A conceptual model of the impact of warming on the active layer of tundra ecosystem processes.** Greenhouse gas pools are represented by green square frames, material pools by yellow square frames, and biological processes by frames in the shape of blue punched tape. Material flows are indicated by thicker black arrows. Impacts of environmental attributes (for example, soil temperature) and microbial community are marked by narrow arrows in black and red, respectively, and labelled with a '+' if increases in gene abundance were observed in this study.

Overall, whether the tundra soil acts as a C source or sink depends on plant and microbial responses to climate warming. Our results indicate that the soil C is highly vulnerable to climate warming and this vulnerability is determined by a set of complex microbial feedbacks to the temperature increase. Improved predictions by ecosystem models to climate warming30 may be possible through better assessment of microbial functional capacities and their responses.

Methods

Site description and sampling.

The Carbon in Permafrost Experimental Heating Research (CiPEHR) site was established in September 2008 at a moist acidic tundra area of Interior Alaska near the Denali National Park in the Eight Mile Lake region (63° 52′ 59″ N, 149° 13′ 32″ W). The experimental plots were located in the discontinuous permafrost region where permafrost thaw has been observed in the past several decades. Experimental design and site description were described in detail previously[14]. Briefly, three experimental blocks were located approximately 100 m away from each other. In each block, two snow fences were erected in the winter of each year (October to April) about 5 m apart. The soil warming treatment plots were located 5 m back from the leeward side of the snow fences, and the paired control plots were at the windward side of the snow fences. Soil temperature was increased in the warmed plots owing to thicker snow cover on the soil surface and lower wind strength. Snow fences were removed in the spring before snow melt to provide uniform hydraulic conditions in both winter warming and control treatments. From 1976 to 2009, mean monthly temperature in the field ranged from −16 °C in December to 15 °C in July, with an annual mean temperature of −1.0 °C. The average annual precipitation was 378 mm. Only $C_3$ plant species were observed in this area. Dominant species include *Eriophorum vaginatum*, *Vaccinium uliginosum*, some other vascular species, non-vascular feather moss and lichen. In the experimental plots, soil from the ground surface to a depth of 45–65 cm, depending on sampling cores, was rich in organic carbon (C) materials; below that depth was mineral soil with a mixture of glacial till and windblown loess. The active-layer depth was about 50 cm.

Twelve soil cores, six from treatment and six from control plots, were taken using electric drills in destructive sampling plots at the six snow fences in the beginning of the 2010 growing season (May), one and one-half years after the initiation of the winter warming treatment. Our analysis provides a snapshot of the soil microbial community response to early stage soil warming. The 15–25 cm depth soil fractions were analysed in this study.

Environmental and soil chemical measurements.

Thaw depth was measured weekly during the growing season (May to September 2010) using a metal depth probe[14]. The thaw depth data presented in this study were the average values for the 2010 growing season.

Constantan–copper thermocouples and CR1000 data loggers (Campbell Scientific) were used to measure and record soil temperature and moisture content at 5, 10, 20 and 40 cm every half hour in flux bases installed in each plot[14]. The soil temperatures in Fig. 1 were reanalysed from previously published data[12], which represent either growing season or wintertime (December 2009 to March 2010) temperatures averaged over 5–40 cm soil depth. To represent the microclimate of the soil where and when the microbial communities were sampled, the soil temperature data used in

canonical correspondence analysis (CCA; Supplementary Fig. 2) were the average values at 20 cm depth from December 2009 to May 2010. Volumetric water content from the soil surface to 15 cm depth was measured using site-calibrated Campbell CS616 water content reflectometer probes[14]. Soil moisture data presented in this study were averaged over the 2010 growing season.

To prepare soils for microbial and chemical analyses, visible roots and stones were removed by metal forceps. To measure soil C and nitrogen (N), soil samples (5 g) were dried at 70 °C until constant weight, ground to powder, encapsulated in silver foil and fumigated with HCl for 24 h at room temperature to remove soil inorganic C (carbonates). Soil C and N concentrations were analysed in the Colorado Plateau Stable Isotope Laboratory at the Northern Arizona University on a DELTA V Advantage isotope ratio mass spectrometer (Thermo Fisher Scientific), configured through a Finnigan CONFLO III (Thermo Fisher Scientific) and using a Carlo Erba NC2100 elemental analyser (CE Elantech). The total organic C (TOC) and soil N content of each sample were calculated as the percentage mass of C or N (ref. 32).

To measure soil C pools, soil samples were processed with a two-step hydrolysis procedure to separate the labile and recalcitrant C pools[24]. First, 5 N $H_2SO_4$ was used to hydrolyse dried soil at 105 °C for 30 min, from which the hydrolysate and wash-offs were collected after centrifugation as labile pool 1, containing mainly polysaccharides. Second, the residue was then shaken continuously overnight at room temperature with 26 N $H_2SO_4$, followed by hydrolysis at 105 °C for 3 h with acid diluted to 2 N. The hydrolysate and wash-offs were recovered as labile pool 2, containing mostly cellulose. The recalcitrant C pool consisted of the remaining organic C. The organic C in labile pools 1 and 2 was analysed using a Shimadzu TOC-V CPH PC-Controlled TOC analyser (Shimadzu Corporation) and the organic C in the recalcitrant C pool was analysed using a PerkinElmer Optima 2000DV ICP-OES spectrometer (PerkinElmer) in the Environmental and Agricultural Testing Service laboratory at North Carolina State University.

Aboveground plant communities.

Aboveground plant community investigations were conducted as described previously[12,14]. In brief, aboveground biomass and net primary productivity (ANPP) were determined by a non-destructive point-frame method using a 60 × 60 cm point frame with a grid size of 8 × 8 cm (ref. 33). At each of the 49 intersecting grid points, a metal rod (1 mm diameter) was placed vertically through the plant canopy. Species identity and tissue type (leaf, stem or fruit) were recorded for every ‘hit’ with the rod. Aboveground live biomass for each vascular plant species, moss and lichen was estimated by applying allometric equations developed for this site to the average number of point-frame ‘hits’ per plot[34]. Vascular plant ANPP was estimated as the sum of the current year’s apical growth (leaves, stems, flowers and fruits) and

secondary growth. The ratio of biomass between each tissue type and total plant was determined from destructive harvesting of a site adjacent to CiPEHR (ref. 35). Secondary growth was evaluated using growth rates determined from tussock tundra at Toolik Lake, Alaska[35]. Moss NPP was measured by the cranked wire method, which measures vertical growth of moss using a stainless-steel reference wire inserted at the moss surface[34,36]. Three to five cranked wires were placed in four moss types in each treatment at all fences to measure the growth from mid-May to mid-September. Feather moss NPP was estimated as the product of linear growth per stem, stem density, biomass per unit stem growth and percentage cover. Allometric equations developed for the Eight Mile Lake watershed[34] as well as percentage cover were used to convert the vertical growth of other types of moss into biomass. Moss NPP was the sum of all types of moss NPP. Current year's fully formed green leaves from six vascular plants found across plots were collected at peak biomass (mid-July) for measuring foliar N and at the end of the growing season (late September) for senescent N (ref. 14). At least three leaves from two to three individuals in each plot were collected each time. Leaves were dried at 60 °C, finely ground, and analysed on a continuous flow isotope ratio mass spectrometer (Thermo Fisher Scientific) coupled with a Costech elemental analyser.

Decomposition.

Weighed cellulose filter papers (Fisher brand P8 09-802-1B) were placed into fibreglass mesh bags and placed vertically at 0–10 cm in the field soils in September 2009 and collected in September 2010. The bags were rinsed and dried at 60 °C for weighing. The percentage of mass loss was calculated to represent decomposition rate.

Ecosystem C flow.

Ecosystem C flux measurements were described previously[14,23]. Growing season net ecosystem exchange (NEE) and ecosystem respiration ($R_{eco}$) were measured from May to September 2010 using an automated $CO_2$ flux system coupled to the flux chambers[14]. $R_{eco}$ was determined with night measurements. Gross primary productivity (GPP) was estimated as the difference between NEE and $R_{eco}$ (values are positive for C flowing from atmosphere to terrain and vice versa). Winter respiration was estimated using a parameterized winter respiration model, adjusted using in-plot winter respiration measurements in March and April 2009 using an infrared gas analyser in a portable $CO_2$ flux system. In winter, there was no photosynthetic activity and $R_{eco}$ represents mainly microbial respiration. The C flux data used for analysis in this study were reanalysed from previous published data sets[23].

Soil DNA extraction.

Soil DNA was extracted using a PowerMax Soil DNA Isolation Kit (MO BIO), and the quality was assessed based of spectrometry absorbance at

wavelengths of 230 nm, 260 nm and 280 nm (ratios of absorbance at 260/280 nm around 1.8, and 260/230 nm > 1.7) detected by a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Then it was quantified with Pico Green using a FLUOstar OPTIMA fluorescence plate reader (BMG LabTec) before being used for gene array labelling and sequencing library preparation. Detailed protocols for soil microbial community analysis (for example, amplicon sequencing, shotgun sequencing and GeoChip) are provided as Supplementary Information.

GeoChip analysis.

GeoChip 4.2 is a comprehensive gene array containing 107,950 probes designed for covering 792 functional gene families from 11 major functional categories including C, N, phosphorus and sulfur cycling[37,38]. One microgram of DNA from each sample was mixed with random primers and denatured before dNTP, fluorescent dye Cy-3 dUTP and DNA polymerase were added for labelling at 37 °C for 6 h, followed by heating at 95 °C for 3 min. Labelled DNA was purified and dried up. For hybridization, DNA was resuspended in hybridization solution containing a sample tracking control, formamide, SSC, SDS, a Cy3-labelled alignment oligonucleotide, a Cy5-labelled alignment oligonucleotide and a Cy5-labelled common oligonucleotide reference standard target. After denaturing, the mixtures were deposited onto the glass microarray and hybridized at 42 °C for 16 h. Then the arrays were washed and dried, and scanned by an MS 200 Microarray Scanner (NimbleGen) at 532 nm and 635 nm. NimbleScan software version 2.5 (NimbleGen) was used to grid and process the images to transform them into signal intensity. The raw signals from NimbleScan were submitted to the Microarray Data Manager on our website (http://ieg.ou.edu/microarray), cleaned, normalized and analysed using the data analysis pipeline. Briefly, spot signal-to-noise ratio and minimum intensity cutoff were used as standard to remove unreliable spots. Both the universal standard and functional gene spot intensities are used to normalize the signals among arrays. Data were log transformed after cleaning and normalization. A total of 48,188 functional gene probes were detected across all samples in this study.

Illumina MiSeq sequencing of 16S rRNA gene amplicons.

DNAs were amplified for the V4 region of 16S rRNA genes using primer set 515F and 806R, and sequenced in one run on a MiSeq using 2 × 150 pair end format[39]. Raw sequences were assembled using RDP's paired-end reads Assembler. Any assembled sequences with any ambiguous bases ('N') were discarded. Then, 5.28% of the remaining reads were identified as chimaeras using Uchime[40] and removed. The remaining sequences were clustered into operational taxonomic units (OTUs) using Uclust[41] at 97% identity, and randomly resampled to the depth of 42,684 reads per sample. Representative sequences chosen by Uclust from each OTU were annotated

taxonomically using the RDP Classifier[42] with the confidence cutoff 0.5. Finally, 512,208 sequences in 23,677 OTUs were obtained.

454 pyrosequencing of *nifH* gene amplicons.

*nifH* genes from the DNA samples were amplified using the primer pair *nifH* Poly F (5′-TGCGAYCCSAARGCBGACTC-3′) and Poly R (5′-ATSGCCATCATYTCRCCGGA-3′) and sequenced on the 454 GSFLX Titanium platform at Macrogen[43,44,45]. After trimming primers, the sequences were cleaned using LUCY (ref. 46). Sequences with 'N', those containing frameshift(s) detected by FrameBot[47], and those identified as chimaeras by Uchime[40] according to the Zehr *nifH* Database[48] were all removed from downstream analyses. The remaining 162,523 sequences were clustered into 2,643 non-singleton OTUs using CD-HIT (ref. 49) at 0.95 identity, which was an arbitrary but strict enough cutoff to identify different species according to previous studies on *nifH* and other N-cycling-related genes[50,51,52]. Finally, 2,643 non-singleton OTUs were normalized to relative abundance (scale all of the sample sequence numbers to the largest one) for statistical analyses. The representative sequence for each OTU was assigned taxonomic information using the FrameBot[47] nearest-neighbour match with an identity cutoff of 0.5.

Shotgun metagenome sequencing analysis.

Each soil metagenome was prepared using the TruSeq Kit and sequenced at the Los Alamos National Laboratory Genome Facility using the Illumina HiSeq 2000 in one flow cell lane with a 2 × 150 bp paired-end kit[39]. A total of 3.24 billion reads were generated from the 12 samples, with both phylogenetic and functional information extracted. After data processing, it was found that one of the samples from the control (C1) did not produce enough useful sequence during shotgun sequencing and thus this sample was removed from all subsequent analyses. For phylogenetic analysis, the metagenome reads were trimmed[53] and searched against representative OTU sequences from GreenGenes[54] using BLAT (ref. 55). Paired reads that both matched GreenGenes reference sequences were identified as 16S reads and were extracted for further analyses. These 16S reads were subsequently searched against the 99%-clustered GreenGenes OTU sequences. The reads were assigned to the taxon that was the lowest common ancestor of the two reads in a pair. The 797,898 reads were assigned to 23,167 OTUs in total. For functional subsystem analysis, 25 million reads were randomly resampled from each sample. Open reading frames were predicted on non-16S encoding reads using FragGeneScan[56]. The translated amino acid sequences were then searched against the M5NR database[57] using BLAT. Reads matching genes incorporated in the SEED database[58] were assigned to the corresponding best-matched subsystem(s). The numbers of assigned reads were taken as a proxy of abundance of the SEED subsystem(s). An approach combining re-sampling techniques, the DESeq package[59], and binomial testing with adjusted $p$ values[60] was then applied to identify significantly

differentially abundant subsystems (pathways) under warming versus control plots, as described previously[61].

Annotating shotgun sequences on the basis of GeoChip genes.

An ecological functional gene-oriented metagenomic analysis pipeline (EcoFun-MAP) has been developed to fish out sequence reads of important environmental functional genes from shotgun metagenome sequence data. EcoFun-MAP is a method designed for annotating metagenomic sequences by comparing them with functional genes used to fabricate GeoChip. In the preparation of the reference databases, keyword queries were submitted to the NCBI (ref. 62) online GenBank for 308 functional genes to retrieve candidate reference sequences, from which 5 to 200 distinct representative sequences from each gene were manually selected functional gene seed sequences (FGSSs). The selected FGSSs were aligned using both global and local algorithms in ClustalW (ref. 63), and the resulting alignments were used as input for another program HMMBUILD (ref. 64) to build both global and local HMMER (ref. 64) models (FGSS-HMM). Next, the candidate reference sequences for each gene were searched back against corresponding FGSS-HMM using HMMSEARCH (ref. 64). The output sequences, termed functional gene reference sequences (FGRSs), were clustered into OTUs for each gene using CD-HIT at the similarity threshold of 95%. In addition, BLAST databases were constructed on the FGRSs with MAKEBLASTDB (ref. 65). To this end, two reference databases involved in the method were established: FGSS-HMM and FGRS-BLAST. For annotation, sequences from HighSeq were resampled to the minimal number of reads in a sample, and were quality trimmed by Btrim[66]. All trimmed nucleotide sequences were translated into protein sequences using FragGeneScan[56]. HMMSEARCH was used for annotating the predicted protein sequences with the FGSS-HMM database, and both global and local model hits were counted as valid results. Also, all FGSS-HMM confirmed sequences were compared together against the FGRS-BLAST database with BLASTN (ref. 65). Only best hits (Rank No. 1 in BLAST results) between probes and sequences were kept as final processing results. The web-based pipeline application of EcoFun-MAP can be accessed with request.

Statistical analysis.

Statistical analyses were carried out using R software version 2.15.1 using the package vegan (v.2.3-2; ref. 67) when not specified. Detrended correspondence analysis[68] was performed to visualize the overall microbial community composition among samples. Three complementary non-parametric multivariate analyses, non-parametric multivariate analysis of variance (Adonis; ref. 69), analysis of similarity (ANOSIM; ref. 70), and the multi-response permutation procedure (MRPP; ref. 71), were used to test the differences in soil microbial communities between warming and control treatments. CCA (ref. 72) was performed to determine the linkage between environmental variables and microbial community composition. For selecting

environmental variables, those containing redundant information were reduced to minimum number, keeping only the variables that had significant impacts. Also, the final sets of variables should have the variance inflation factors all < 20. Finally, soil temperature, soil moisture and GPP remained in the CCA model (Supplementary Fig. 2a) of GeoChip data. Labile C pool 1 and 2 (%), and soil N content (%) were selected for 16S rRNA gene-based analysis (Supplementary Fig. 2b). The significance of the CCA model was tested by analysis of variance (ANOVA). According to CCA results, variation partitioning analysis was performed to determine the contribution of each individual variable or groups of variables to total variations in soil microbial community compositions. CCA was also used to determine correlations between abundance of subcategories of functional genes and the individual environmental variables (Supplementary Table 5). Two-tailed *t*-tests were performed to examine whether the differences between warming and control treatments were significant on the basis of several important biotic and abiotic variables (that is, soil C contents, aboveground biomass and total bacteria, archaea and fungi abundance) using Microsoft Excel 2010. ANOVA (ref. 73) was performed to test the treatment effect on the abundance of each functional gene involved in C and N cycling for GeoChip or relative abundances of OTUs of certain genus or phylum groups. In addition to the warming treatment effect, the probe or OTU also factored into the model for partitioning the variance of probes within each functional gene. Response ratio was used to compute the effects of warming on functional genes relevant to GeoChip probes from shotgun sequences using the formula described in ref. 74.

Data availability.

Raw shotgun metagenome, 16S rRNA and *nifH* amplicon gene sequences are available in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under study no. PRJEB10725. GeoChip raw and normalized signal intensities can be accessed through the URL ftp://129.15.40.240:8187/nclimate2940/Raw_GeoChip_Data.txt and ftp://129.15.40.240:8187/nclimate2940/Normalized_GeoChip_Data.csv.

References

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.

Schuur, E. *et al.* Expert assessment of vulnerability of permafrost carbon to climate change. *Climatic Change* 119, 359–374 (2013).

3.

Zhou, J. *et al.* Microbial mediation of carbon-cycle feedbacks to climate warming. *Nature Clim. Change* 2, 106–110 (2012).

4.

Graham, D. E. *et al.* Microbes in thawing permafrost: the unknown variable in the climate change equation. *ISME J.* 6, 709–712 (2012).

5.

Lee, H., Schuur, E. A. G., Inglett, K. S., Lavoie, M. & Chanton, J. P. The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. *Glob. Change Biol.* 18, 515–527 (2012).

6.

Hicks Pries, C. E., Schuur, E. A. G. & Crummer, K. G. Holocene carbon stocks and carbon accumulation rates altered in soils undergoing permafrost thaw. *Ecosystems* 15, 162–173 (2012).

7.

Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* 23, GB2023 (2009).

8.

Grosse, G. *et al.* Vulnerability of high-latitude soil organic carbon in North America to disturbance. *J. Geophys. Res.* 116, G00K06 (2011).

9.

Hassol, S. J. *Impacts of A Warming Arctic–Arctic Climate Impact Assessment* (Cambridge Univ. Press, 2004).

10.

Schuur, E. A. G. & Abbott, B. Climate change: high risk of permafrost thaw. *Nature* 480, 32–33 (2011).

11.

Lawrence, D. M., Slater, A. G. & Swenson, S. C. Simulation of present-day and future permafrost and seasonally frozen ground conditions in CCSM4. *J. Clim.* 25, 2207–2225 (2012).

12.

Natali, S. M., Schuur, E. A. G. & Rubin, R. L. Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost. *J. Ecol.* 100, 488–498 (2012).

13.

Walker, M. D. *et al.* Plant community responses to experimental warming across the tundra biome. *Proc. Natl Acad. Sci. USA* 103, 1342–1346 (2006).

14.

Natali, S. M. *et al.* Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra. *Glob. Change Biol.* 17, 1394–1407 (2011).

15.

Yergeau, E. *et al.* Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *ISME J.* 6, 692–702 (2012).

16.

Mackelprang, R. *et al.* Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371 (2011).

17.

Hultman, J. *et al.* Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212 (2015).

18.

Coolen, M. J. L. & Orsi, W. D. The transcriptional response of microbial communities in thawing Alaskan permafrost soils. *Front. Microbiol.* 6, 197 (2015).

19.

Sturm, M., Racine, C. & Tape, K. Climate change: increasing shrub abundance in the Arctic. *Nature* 411, 546–547 (2001).

20.

Walker, D. A. *et al.* The circumpolar arctic vegetation map. *J. Veg. Sci.* 16, 267–282 (2005).

21.

Schuur, E. A. G. *et al.* The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* 459, 556–559 (2009).

22.

Zhao, M. *et al.* Microbial mediation of biogeochemical cycles revealed by simulation of global changes with soil transplant and cropping. *ISME J.* 8, 2045–2055 (2014).

23.

Natali, S. M., Schuur, E. A. G., Webb, E. E., Hicks Pries, C. E. & Crummer, K. G. Permafrost degradation stimulates carbon loss from experimentally warmed tundra. *Ecology* 95, 602–608 (2014).

24.

Rovira, P. & Vallejo, V. R. Labile and recalcitrant pools of carbon and nitrogen in organic matter decomposing at different depths in soil: an acid hydrolysis approach. *Geoderma* 107, 109–141 (2002).

25.

Zhou, J. *et al.* High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio* 6, e02288-14 (2015).

26.

Lau, M. C. Y. *et al.* An active atmospheric methane sink in high Arctic mineral cryosols. *ISME J.* 9, 1880–1891 (2015).

27.

Natali, S. M. *et al.* Permafrost thaw and soil moisture driving $CO_2$ and $CH_4$ release from upland tundra. *J. Geophys. Res.* 120, 525–537 (2015).

28.

IPCC *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).

29.

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

30.

Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J. Clim.* 19, 3337–3353 (2006).

31.

Hicks Pries, C. E., Schuur, E. A. G., Vogel, J. G. & Natali, S. M. Moisture drives surface decomposition in thawing tundra. *J. Geophys. Res.* 118, 1133–1143 (2013).

32.

Avramidis, P., Nikolaou, K. & Bekiari, V. Total organic carbon and total nitrogen in sediments and soils: a comparison of the wet oxidation–titration method with the combustion-infrared method. *Agric. Agric. Sci. Procedia* 4, 425–430 (2015).

33.

Walker, M. Community baseline measurements for ITEX studies. *ITEX Manual* 2, 39–41 (1996).

34.

Schuur, E. A. G., Crummer, K., Vogel, J. & Mack, M. Plant species composition and productivity following permafrost thaw and thermokarst in Alaskan tundra. *Ecosystems* 10, 280–292 (2007).

35.

Shaver, G. R. *et al.* Species composition interacts with fertilizer to control long-term change in tundra productivity. *Ecology* 82, 3163–3181 (2001).

36.

Clymo, R. S. The growth of Sphagnum: methods of measurement. *J. Ecol.* 58, 13–49 (1970).

37.

Tu, Q. *et al.* GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Mol. Ecol. Resour.* 14, 1755-0998.12239 (2014).

38.

Wu, L. *et al.* Phasing amplicon sequencing on Illumina MiSeq for robust environmental microbial community analysis. *BMC Microbiol.* 15, s12866-015-0450-4 (2015).

39.

He, Z. *et al.* GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* 1, 67–77 (2007).

40.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200 (2011).

41.

Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461 (2010).

42.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267 (2007).

43.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. & Welch, D. M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8, R143 (2007).

44.

Meyer, M., Stenzel, U. & Hofreiter, M. Parallel tagged sequencing on the 454 platform. *Nature Protoc.* 3, 267–278 (2008).

45.

Ronaghi, M., Uhlén, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* 281, 363–365 (1998).

46.

Chou, H. H. & Holmes, M. H. DNA sequence quality trimming and vector removal. *Bioinformatics* 17, 1093–1104 (2001).

47.

Wang, Q. *et al.* Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio* 4, e00592-13 (2013).

48.

Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ. Microbiol.* 5, 539–554 (2003).

49.

Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659 (2006).

50.

Palmer, K., Drake, H. L. & Horn, M. A. Genome-derived criteria for assigning environmental *narG* and *nosZ* sequences to operational taxonomic units of nitrate reducers. *Appl. Environ. Microbiol.* 75, 5170–5174 (2009).

51.

Mao, Y., Yannarell, A. C. & Mackie, R. I. Changes in N-transforming archaea and bacteria in soil during the establishment of bioenergy crops. *PLoS ONE* 6, e24750 (2011).

52.

Pereira e Silva, M. C., Schloter-Hai, B., Schloter, M., van Elsas, J. D. & Salles, J. F. Temporal dynamics of abundance and composition of nitrogen-fixing communities across agricultural soils. *PLoS ONE* 8, e74500 (2013).

53.

Luo, C., Tsementzi, D., Kyrpides, N. C. & Konstantinidis, K. T. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J.* 6, 898–901 (2011).

54.

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).

55.

Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002).

56.

Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191 (2010).

57.

Wilke, A. *et al.* The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* 13, 1471-2105-13-141 (2012).

58.

Overbeek, R. *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702 (2005).

59.

Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* 11, R106 (2010).

60.

Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 289–300 (1995).

61.

Luo, C., Rodriguez, R. L. & Konstantinidis, K. T. A user’s guide to quantitative and comparative analysis of metagenomic datasets. *Methods Enzymol.* 531, 525–547 (2013).

62.

Geer, L. Y. *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* 38, D492–D496 (2010).

63.

Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948 (2007).

64.

Eddy, S. R. Profile hidden Markov models. *Bioinformatics* 14, 755–763 (1998).

65.

Camacho, C. *et al.* BLAST + : architecture and applications. *BMC Bioinformatics* 10, 1471-2105-10-421 (2009).

66.

Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 98, 152–153 (2011).

67.

Oksanen, J. *et al. vegan: Community Ecology Package* R package version 2.3-2 (R Foundation, 2015).

68.

Oksanen, J. & Minchin, P. R. Instability of ordination results under changes in input data order: explanations and remedies. *J. Veg. Sci.* 8, 447–454 (1997).

69.

Zapala, M. A. & Schork, N. J. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc. Natl Acad. Sci. USA* 103, 19430–19435 (2006).

70.

Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143 (1993).

71.

Sickle, J. V. Using mean similarity dendrograms to evaluate classifications. *J. Agric. Biol. Environ. Stat.* 2, 370–388 (1997).

72.

Hotelling, H. in *Breakthroughs in Statistics* (eds Kotz, S. & Johnson, N.) 162–190 (Springer, 1992).

73.

Chambers, J., Freeny, A. & Heiberger, R. in *Statistical Models in S* (eds Chambers, J. M. & Hastie, T. J.) 145–193 (Wadsworth Brooks/Cole, 1992).

74.

Luo, Y., Hui, D. & Zhang, D. Elevated $CO_2$ stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. *Ecology* 87, 53–63 (2006).

Acknowledgments

This material is based upon work supported by the US Department of Energy, Office of Science, Genomic Science Program under Award Numbers DE-SC0004601 and DE-SC0010715, the NSF LTER program, the Office of the Vice President for Research at the University of Oklahoma, and the Collaborative Innovation Center for Regional Environmental Quality.