

# Lawrence Berkeley National Laboratory

## Recent Work

### Title

Landscape topography structures the soil microbiome in arctic polygonal tundra.

### Permalink

<https://escholarship.org/uc/item/9tk1s8qz>

### Journal

Nature communications, 9(1)

### ISSN

2041-1723

### Authors

Taş, Neslihan  
Prestat, Emmanuel  
Wang, Shi  
et al.

### Publication Date

2018-02-01

### DOI

10.1038/s41467-018-03089-z

Peer reviewed

ARTICLE

DOI: 10.1038/s41467-018-03089-z

OPEN

# Landscape topography structures the soil microbiome in arctic polygonal tundra

Neslihan Taş<sup>1</sup>, Emmanuel Prestat<sup>1,2</sup>, Shi Wang<sup>1</sup>, Yuxin Wu<sup>1</sup>, Craig Ulrich<sup>1</sup>, Timothy Kneafsey<sup>1</sup>, Susannah G. Tringe<sup>3</sup>, Margaret S. Torn<sup>1,4</sup>, Susan S. Hubbard<sup>1</sup> & Janet K. Jansson<sup>5</sup>

In the Arctic, environmental factors governing microbial degradation of soil carbon (C) in active layer and permafrost are poorly understood. Here we determined the functional potential of soil microbiomes horizontally and vertically across a cryoperturbed polygonal landscape in Alaska. With comparative metagenomics, genome binning of novel microbes, and gas flux measurements we show that microbial greenhouse gas (GHG) production is strongly correlated to landscape topography. Active layer and permafrost harbor contrasting microbiomes, with increasing amounts of *Actinobacteria* correlating with decreasing soil C in permafrost. While microbial functions such as fermentation and methanogenesis were dominant in wetter polygons, in drier polygons genes for C mineralization and CH<sub>4</sub> oxidation were abundant. The active layer microbiome was poised to assimilate N and not to release N<sub>2</sub>O, reflecting low N<sub>2</sub>O flux measurements. These results provide mechanistic links of microbial metabolism to GHG fluxes that are needed for the refinement of model predictions.

<sup>1</sup>Earth and Environmental Sciences Area Lawrence Berkeley National Laboratory, One Cyclotron Road, MS B74, Berkeley, CA 94720, USA. <sup>2</sup>HaliuDx, Luminy Biotech Entreprises, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France. <sup>3</sup>DOE Joint Genome Institute (DOE-JGI), 2800 Mitchell Dr, Walnut Creek, CA 94598, USA. <sup>4</sup>Energy and Resources Group, University of California, Berkeley, California 94720, USA. <sup>5</sup>Biological Sciences Division Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA. Correspondence and requests for materials should be addressed to N.T. (email: [ntas@lbl.gov](mailto:ntas@lbl.gov)) or to J.K.J. (email: [janet.jansson@pnnl.gov](mailto:janet.jansson@pnnl.gov))

Ice wedge polygons are important topographical and hydrological features in otherwise low relief coastal arctic tundra<sup>1</sup>. Ice wedge development begins at the ground surface as frost cracks and forms a polygonal pattern in the top soil as ice wedges enlarge over time<sup>2</sup>. Approximately 20% of the Arctic Coastal Plain of northern Alaska contains polygonal grounds and thaw lakes that develop in ice-rich permafrost<sup>3</sup>. Depending on their growth or degradation state, ice wedges can create various sizes (~5–20 m) and types of polygons. Development of this microtopography regulates water distribution over large areas and effects soil properties that are important for C cycling, such as soil moisture, other biogeochemical cycles, and freeze-thaw dynamics<sup>2</sup>. With a predicted increase in surface air temperature of 5–6 °C by the end of this century<sup>4</sup> an accelerated soil warming<sup>5</sup> in the polygonal tundra will lead to a deepening of the seasonally-thawed active layer, enlarge permafrost thaw area and changes in the distribution of water across this landscape. Therefore, the estimated 1035 Pg C<sup>6</sup> stored in arctic ecosystems could undergo rapid microbial decomposition, resulting in a significant positive feedback on the climate in the form of greenhouse gas (GHG) emissions<sup>7</sup>.

A major challenge in linking microbial functions to GHG emissions in arctic soils is the identification of environmental factors that govern the distribution of microbial functions and C flux<sup>7</sup>. Besides the complexity introduced by the high microbial diversity in these soils<sup>7</sup>, our understanding of environmental controls and spatial variation on microbial C metabolism, with respect to hydrological patterns<sup>8</sup>, is limited. To resolve the microbial metabolism leading to GHG emissions across the polygonal tundra in the Arctic, we used multiple DNA sequencing approaches to determine the soil microbial community composition and metabolic potential. We focused our studies in the Barrow Environmental Observatory (BEO) located in Barrow, Alaska. Polygonal grounds constitute 65% of the surface in the Barrow Peninsula<sup>3</sup>. The landscape of the BEO includes a mosaic of drained thaw lakes and interstitial polygons where different polygon types—high-centered (HC), flat-centered (FC) and low-centered (LC)—represent different states of permafrost degradation and moisture distribution<sup>1</sup> (Fig. 1). Variance in landscape features (e.g., rim, center, and trough) have been hypothesized to result in distinct microbial habitats that might respond differently to warming<sup>2</sup>. We collected 29 active layer soil cores along a 500 m transect that extended laterally across HC, FC, and LC polygons in the BEO in September 2011. The top 0–50 cm, including surface organic (0–10 cm) and deeper mineral (20–50 cm) layers of the active layer, was collected from centers, rims, and troughs of the polygons along the transect<sup>1</sup> (Fig. 1). Additionally, two deep cores, 1 and 2.65 m in depth, were taken from a FC polygon within the transect in April 2012 to access the vertical distribution of the microbial communities in the permafrost. The soil microbiome compositions in the samples were determined by sequencing of 16S rRNA genes (16S) and the functional potential was inferred by sequencing of total metagenomic DNA. The sequence data were correlated to in situ GHG flux measurements and geophysical and geochemical soil characteristics<sup>1</sup>. Here we show that soil microbiomes and their functional potentials in arctic polygonal grounds are governed by the landscape microtopography.

## Results

### Microbial community composition is linked to polygon types.

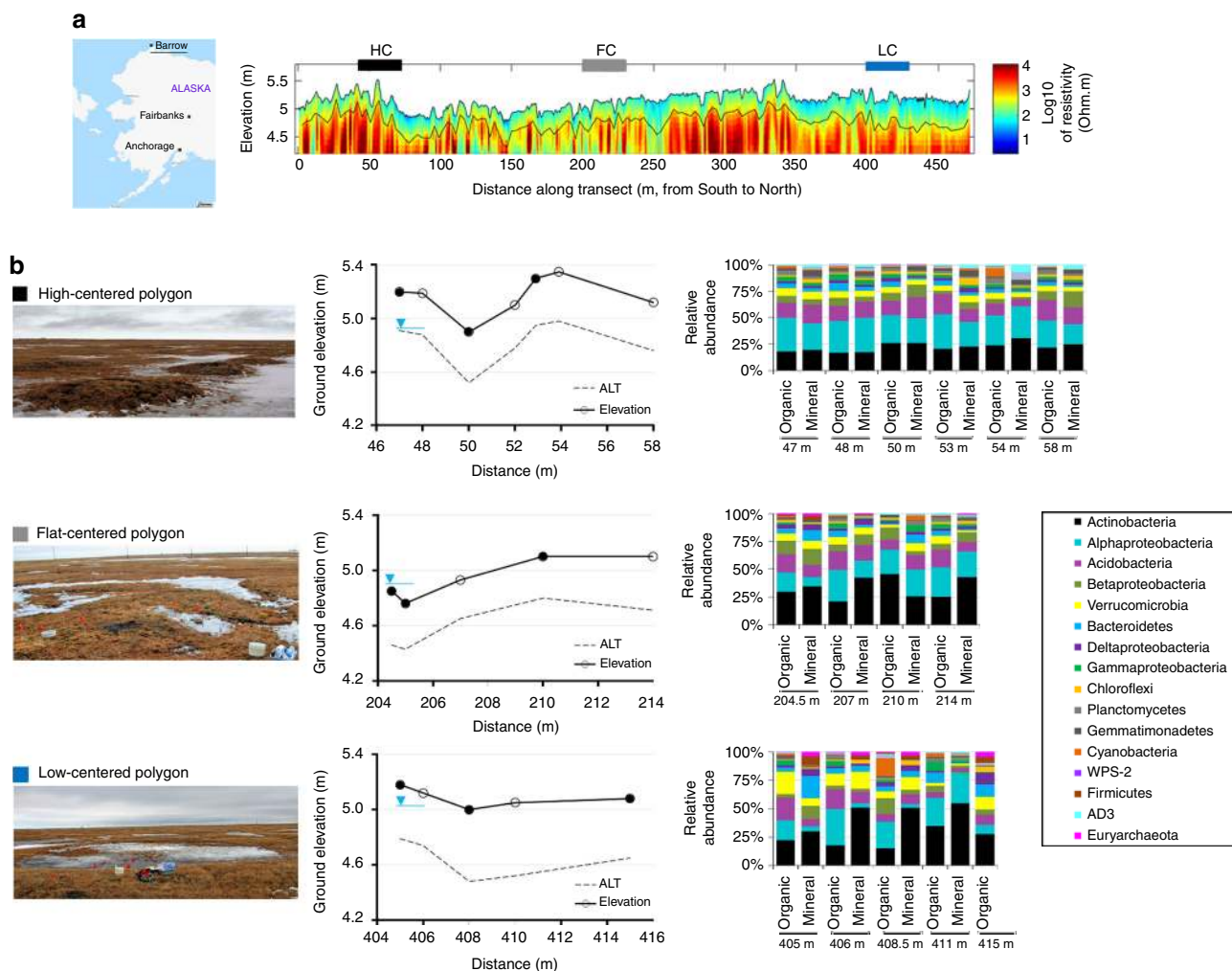
We found a shift in microbial community composition along the polygon transect, with different polygon types and soil horizons having distinct microbiomes (Fig. 1, Supplementary Fig. 1). Although the soil microbiomes primarily clustered by polygon

type (ANOSIM  $R = 0.432$ ,  $p < 0.001$ ) (Supplementary Fig. 1), the two soil horizons were populated with phylogenetically different microbial populations (organic and mineral; ANOSIM  $R = 0.117$ ,  $p = 0.044$ ). We detected a weak spatial autocorrelation between different polygon types (Moran's I: 0.144, SD: 0.018,  $p = 1.964E-05$ ) but not between soil horizons (Moran's I: -0.035, SD: 0.079,  $p = 0.983$ ). Additionally, there were significant correlations between organic C content (Adonis analysis:  $F_{1,28} = 2.02$ ,  $r^2 = 0.057$ ,  $p = 0.032$ ) and total nitrogen (N) content ( $F_{1,28} = 2.33$ ,  $r^2 = 0.065$ ,  $p = 0.019$ ) of the samples and the observed microbial community clustering patterns (Supplementary Fig. 1). These results are in alignment with parallel measurements of the soils' geophysical and geochemical properties, which demonstrated differences among polygon types<sup>2</sup>. The phylogenetic alpha diversity (Faith's PD) was also significantly different among the different polygons (ANOVA  $F_{2,27} = 18.64$ ,  $p < 0.001$ ) and soil horizons (ANOVA  $F_{1,28} = 14.34$ ,  $p < 0.001$ ) (Supplementary Table 3, Supplementary Fig. 2). However, within each individual polygon type no significant differences were observed between the alpha diversity of organic and mineral horizons. In general, the HC polygons (both horizons) had the highest alpha diversity (Supplementary Fig. 2).

*Proteobacteria* (HC  $40 \pm 10\%$ , FC  $35 \pm 14\%$ , LC  $28 \pm 22\%$ ) and *Actinobacteria* (HC  $22 \pm 5\%$ , FC  $34 \pm 10\%$ , LC  $35 \pm 15\%$ ) were the two most abundant microbial phyla across the transect (Fig. 1). Using canonical correspondence analysis, we found that 55% of the observed variation in microbial community composition at the OTU level could be explained by a combination of environmental features (Supplementary Fig. 3), including polygon type, soil horizon, pH, total C, and organic C content. HC and FC polygons were dominated by *Alphaproteobacteria* (ANOVA  $F_{2,27} = 3.64$ ,  $p = 0.015$ ) and *Acidobacteria* (ANOVA  $F_{2,27} = 6.31$ ,  $p = 0.016$ ) whereas the wetter LC polygons contained significantly more *Bacteroidetes* (ANOVA  $F_{2,27} = 5.15$ ,  $p = 0.031$ ) and *Verrucomicrobia* ( $F_{2,27} = 7.22$ ,  $p = 0.012$ ) (Fig. 1, Supplementary Fig. 3). In addition, the deeper mineral soil horizons had higher relative levels of *Actinobacteria* (ANOVA  $F_{1,28} = 6.00$ ,  $p = 0.021$ ) (HC  $24 \pm 5\%$ , FC  $37 \pm 9\%$ , LC  $47 \pm 11\%$ ) than the surface organic soil horizons (Fig. 1).

Several functional traits were derived for representative members of the soil microbiomes across the polygon transect based on the similarities of 16S sequences to known species with characterized physiologies. Representatives of methanogenic *Euryarchaeota* were more abundant in the LC polygons (max. 4.6% relative abundance), with high levels of *Methanobacterium* and *Methanosaeta* in the organic horizons and *Methanosaeta* and *Methanosarcina* in the mineral horizons (Supplementary Fig. 4a). Many different representatives of potential methane oxidizing Bacteria were also detected in all of the polygons, including *Verrucomicrobia* (0.02–0.37%) and *Alpha-* (0.04–0.92%, most abundant genus: *Methylocella*), *Beta-* (0.12–1.00%, most abundant genus: *Methylibium*) and *Gammaproteobacteria* (0.01–0.28%, most abundant genus: *Crenothrix*) (Supplementary Fig. 4b). The organic horizons of the LC polygons had a higher relative abundance of methylotrophic bacteria (max. 2.5% relative abundance) in comparison to mineral layers. On the other hand, representative ammonia-oxidizers were more prominent in the HC polygons, including *Candidatus Nitrososphaera*, *Nitrospira*, and *Nitrobacter* with relative abundances varying between 0.08–0.15% (Supplementary Fig. 4c).

In addition to sampling along the lateral transect, we also sampled two deep cores (1 and 2.65 m depth) from a FC polygon to determine the microbial community distribution with depth into the permafrost (Fig. 2). We analyzed the geochemistry and the soil microbiome in 5 cm intervals through these cores (Supplementary Table 1). While organic and mineral layers were

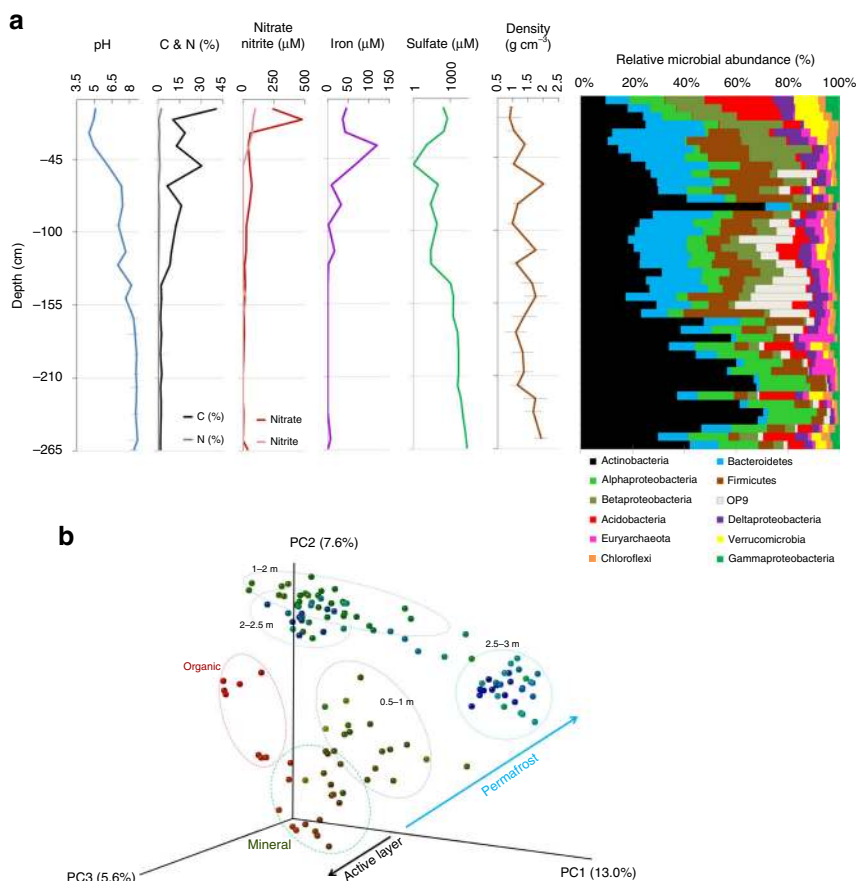


**Fig. 1** Microbial communities of active layer are strongly correlated to landscape topography in arctic polygonal tundra. Samples were collected from active layer soils and permafrost layer along a transect of high- (black), flat- (gray) and low- (blue) centered polygons located at Barrow Experimental Observatory. Map is adapted from ©OpenStreetMap contributors, licensed CC-BY-SA. **a** Electrical resistivity tomography (ERT) data were collected along the ~480 m transect, coincident with soil core retrieval and many different types of in situ electrical resistivity measurements. ERT data were used to characterize deeper permafrost variability and ice-wedge structures (deeper yellow-red-blue), as well as active layer variability (blue-green). Along this ERT transect, the first 0–150 m were dominated by HC polygons (black bar) which transitioned to FC (gray bar) and LC (blue bar) polygons afterwards. ERT and soil characterization data are described elsewhere<sup>1</sup>. **b** Photographs (taken by the authors) show the differences in surface soil morphology among different polygon types. In HC polygons centers and troughs could have an elevation difference up to 0.6 m whereas elevation difference among rims, troughs, and centers of FC and LC polygons vary between 0.1–0.3 m. We collected samples for sequencing of the microbial community composition along the polygonal transect (circles show the sampling locations). Active layer thickness (ALT) was also measured at each sampling point. Water table (blue upside down triangle) levels are inferred from measured water levels in troughs and soil moisture measurements and show an estimated depth. CO<sub>2</sub> and CH<sub>4</sub> fluxes were measured in two consecutive years, 2012 and 2013 from rims, troughs, and centers of polygons (closed circles; Supplementary Fig. 9 and Supplementary Fig. 10)

acidic, soil pH increased up to 8.5 with depth (Fig. 2, Supplementary Table 1). Soil C content on the other hand, decreased with depth. NO<sub>3</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup>, and total Fe concentrations also decreased with depth whereas SO<sub>4</sub><sup>-</sup> concentrations increased in the deep permafrost layers (Fig. 2). *Actinobacteria* was the most abundant phylum throughout the soil profiles, accounting up to 68% of the total microbiome. Within this phylum, *Actinomycetales* and *Solirubrobacterales* were the most abundant orders at 42% and 30% relative abundance, respectively (Supplementary Fig. 5). Although *Actinomycetales* were abundant along the whole depth profile, the relative abundance of *Solirubrobacterales* increased in deeper permafrost layers (1–2.64 m). Most of the *Actinobacterial* orders, except *Acidimicrobiales*, were negatively correlated to high C, N, total Fe, NO<sub>3</sub><sup>-</sup>, and NO<sub>2</sub><sup>-</sup> content in the organic active layer soils. However,

below the active layer, correlations between *Actinobacterial* orders and sample chemistry were marginal. The permafrost layers also had high relative amounts of *Bacteroidetes* (max. 29%) and candidate phylum OP9 (max. 23%). Additionally, more *Euryarchaeota* were detected in the deeper permafrost samples compared to the active layer (max. 13% and max. 3% relative abundance, respectively).

The microbial diversity of the active layer was significantly different from that of the permafrost (Supplementary Fig. 6). Alpha diversity (Faith’s PD and Shannon H’) generally decreased with depth, but some depths had transient spikes in diversity (Supplementary Fig. 5). We also found that the diversity decreased in the transitional layers between the active layer and permafrost, at ~50 cm depth. At this depth we detected higher pH and C content in comparison to the overlying acidic active layer.



**Fig. 2** Active layer and permafrost have contrasting microbiomes. Duplicate deep cores from a FC polygon were dissected in 5 cm intervals and analyzed to determine chemical and microbial composition. **a** Changes in soil and permafrost chemistry and microbial community composition with depth are reported as averages ( $n = 8$  between 0 to 1 m,  $n = 4$  between 1 to 2.64 m). Standard deviations of chemical measurements are represented by gray bars. Standard deviations among technical and biological replicates of microbial community relative abundances are provided in Supplementary Table 5. **b** Unifrac analysis of 16S rRNA sequence variation along the depth gradient from duplicate cores of a FC polygon. Samples were grouped according to organic (red) and mineral (brown-green) in active layer and permafrost (green-blue) horizons. The variance explained by each principal component (PC) axis is given in parentheses

We hypothesize that the abrupt shift in soil chemistry resulted in the decreased diversity at this location.

The microbiomes also differed in composition with depth (Fig. 2). The observed clustering was strongly and significantly influenced by pH, depth and changes in material density (calculated from CT-scanning images; Fig. 2, Table 1). Network analyses of the ratio of shared total operational taxonomic units (OTUs), revealed differences in microbiome clustering patterns with depth (Supplementary Fig. 7). Interestingly, no clusters were shared between organic and mineral horizons of the active layer samples. Permafrost samples on the other hand showed stronger clustering patterns with increasing depth, emphasizing the convergence to more similar microbiomes with depth.

**Microbial genomic potential for GHG production.** The metagenome data revealed both phylogenetic and functional variations across the polygon transect (ANOVA  $F_{2,14} = 1.79$ ,  $p = 0.047$ ) (Fig. 3). The functional gene repertoires were based on functional assignments of the metagenomes using the FOAM<sup>9</sup> database. FOAM enables screening of metagenomes for environmentally relevant genes, including global C and N cycles, derived from KEGG Orthologs (KOs). The functional gene compositions of the metagenomes from HC polygons were highly similar across the organic and mineral layers. By contrast there were differences in

functional gene compositions between the layers for the FC and LC polygons. Specific high-level FOAM<sup>9</sup> categories (Supplementary Fig. 8) that differed significantly across the polygon transect included methanogenesis (ANOVA  $F_{2,14} = 5.34$ ,  $p = 0.012$ ) and transporters (ANOVA  $F_{2,14} = 4.018$ ,  $p = 0.030$ ). Genes for nitrification,  $\text{NO}_3^-$  reduction and assimilation, stress response and pyrimidine metabolism pathways were more abundant in the HC polygons (Fig. 3a), whereas genes for  $\text{H}_2$  production, ABC transporters, valine, leucine, and isoleucine degradation and biosynthesis pathways were more abundant in the FC polygons. The LC polygons had higher amounts of genes for fermentation, N-fixation, osmotic stress response, glycosidase hydrolase, and glutamine/glutamate metabolism.

Previously, significant seasonal differences were observed across the polygon transect, such as higher  $\text{CH}_4$  flux from the LC polygons in early summer compared to late summer<sup>10</sup>. Here we identified subunits of a key enzyme responsible for  $\text{CH}_4$  production (methyl coenzyme M reductase, *mcrABG*) that was significantly (ANOVA  $F_{2,14} = 3.41$ ,  $p = 0.045$ ) higher in the LC polygons compared to the other polygon types (Fig. 3b), corresponding to higher abundances of methanogen 16S rRNA genes in the LC polygons (Fig. 3b). Active layer thickness (ALT) was also significantly deeper (ANOVA  $F_{2,27} = 8.14$ ,  $p = 0.004$ ) in this polygon type due to the more extensive thaw. Subsequent sampling of the site revealed that the LC polygons were consistent

**Table 1 Significant contributions of soil geochemistry to observed differences in prokaryotic diversity in permafrost were tested using ANOVA (df = 21)**

	F Model	p
Depth	4.3	0.002
pH	4.3	0.004
Density	3.8	0.009
Nitrate	4.1	0.013
Total iron	3.3	0.015
Total C	2.7	0.040
Cl	2.2	0.084
Al	1.0	0.431
Na	1.0	0.439
Nitrite	0.9	0.457
Total N	1.0	0.462
Mg	0.7	0.583
C/N	0.6	0.762
Ca	0.5	0.832
K	0.4	0.888
Sulfate	0.4	0.906

sources of CH<sub>4</sub> (Supplementary Fig. 9, Supplementary Fig. 10). The in situ CH<sub>4</sub> flux was significantly higher (Akaike Information Criterion (AIC): 301.04,  $F_{2,43} = 11.17$ ,  $p = 0.00013$ ) in centers of LC polygons compared to the other polygon types (Fig. 3b), although CH<sub>4</sub> fluxes were sporadically detected at lower rates in wetter areas, such as troughs, along the transect (Supplementary Fig. 9, Supplementary Fig. 10). Both phylogenetic marker gene and functional gene distributions in the LC polygons suggest that multiple Archaeal species and pathways were involved in the CH<sub>4</sub> production, including genes from hydrogenotrophic (CO<sub>2</sub>-reducing), acetoclastic, and methylotrophic methanogens, with no significant differences in their relative distributions (Supplementary Fig. 11) in organic and mineral layers. LC polygons had 2-fold higher relative abundance of methanogenesis genes in comparison to HC and FC polygons (Supplementary Fig. 11). By contrast, the potential for CH<sub>4</sub> oxidation was detected along the entire transect. This result is noteworthy, as it shows the co-occurrence of aerobic and anaerobic populations and metabolisms. The relative abundances of particulate methane monooxygenase (*pmoABC*), soluble methane monooxygenase (*mmoXYZ*), and methanol dehydrogenase (*mxhFJGD*) genes were mostly similar between polygon types and soil horizons (Fig. 3b). While HC polygon samples were largely (>50%) populated by Type II methanotrophs, we found similar distributions of Type I and II methanotrophs in FC and LC polygons (Supplementary Fig. 12). There were significantly more genes for oxygen stress response in the HC polygons compared to the LC polygons (Supplementary Fig. 22, ANOVA  $F_{2,14} = 18.90$ ,  $p = 0.023$ ), reflecting the drier and more aerobic conditions in the HC polygons. In particular, the “Cellular response to stress” group that contains genes encoding peroxidases, catalase, and superoxide dismutase was abundant in the HC polygons. These genes are found in aerobes and facultative anaerobes that express them in response to O<sub>2</sub> exposure<sup>9</sup>.

The metagenome predictions were also corroborated by measured rates of CO<sub>2</sub> flux, which were significantly higher in HC polygons ( $p = 0.015$ ) (Fig. 3b); we did not observe any significant difference between CO<sub>2</sub> flux of FC and LC polygons ( $p = 0.055$ ). However strong seasonal variations (measured from July to October) in CO<sub>2</sub> fluxes from the BEO were also observed (Supplementary Fig. 9) (AIC:175.0,  $F_{3,61} = 28.95$ ,  $p = 5.997 \times 10^{-12}$ ). CO<sub>2</sub> fluxes were highest in summer months (July and August) in

all polygon types<sup>10</sup> (Supplementary Fig. 9). Predictions of the microbial metabolic routes leading to CO<sub>2</sub> production included the relative amounts of genes encoding carbohydrate-active enzymes (CAZymes<sup>11</sup>). In the HC polygons there were several genes encoding enzymes for degradation of complex C polymers that were significantly and up to a fold more abundant when compared to the other polygon types. These included: xylan 1,4-beta-xylosidase (*xynB*, EC:3.2.1.37), and bifunctional chitinase/lysozyme (*chiA*, EC:3.2.1.14, EC:3.2.1.17) (ANOVA *xynB*  $F_{2,14} = 4.00$ ,  $p = 0.025$ ; *chiA*  $F_{2,14} = 2.60$ ,  $p = 0.045$ ) (Supplementary Fig. 13). By contrast, genes encoding enzymes for metabolism of oligosaccharides and other simple carbohydrates were more predominant in the LC polygons, including alpha-mannosidase (EC:3.2.1.24) (ANOVA  $F_{2,14} = 3.2$ ,  $p = 0.049$ ), sugar utilization via hexose degradation (ANOVA  $F_{2,14} = 4.18$ ,  $p = 0.038$ ), starch degradation (Supplementary Fig. 14, Supplementary Fig. 15) and ABC transporters involved in sugar transport (Supplementary Fig. 23). The LC polygons also had a significantly higher proportion of genes involved in mixed acid fermentation pathways (Spearman's rho = 0.44,  $p = 0.041$ ) compared to the other polygon types. However, there were no significant differences in many other central metabolic pathways among polygons, including pyruvate fermentation (Supplementary Fig. 16), homoacetogenesis, fatty acid oxidation, or hydrocarbon degradation (Supplementary Fig. 17, Supplementary Fig. 18, Supplementary Fig. 19).

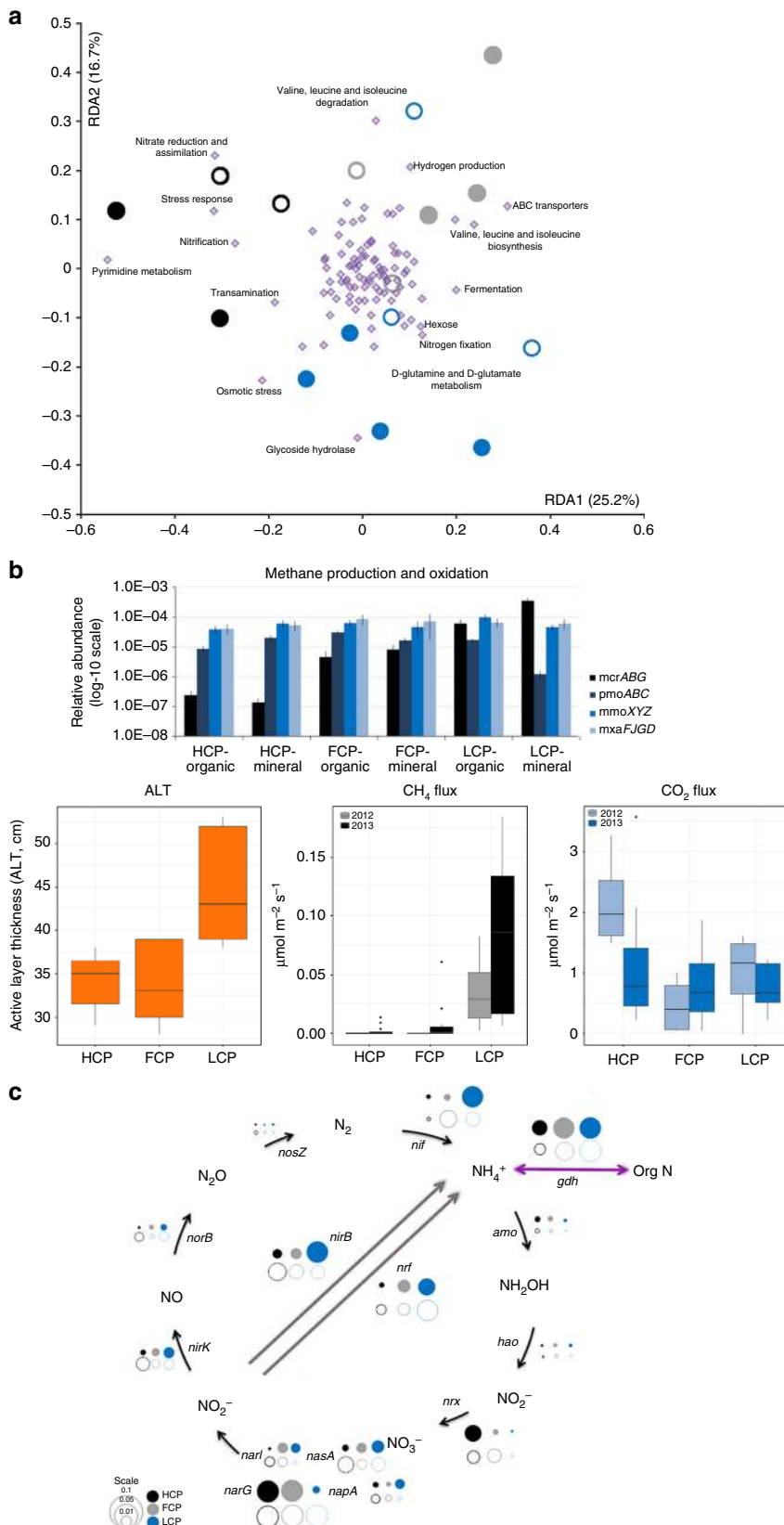
Several genes encoding anaerobic terminal electron acceptor pathways were represented in the metagenomes. The LC polygons had up to 2-fold higher levels of decaheme cytochromes (*mtr*) (Supplementary Fig. 20), suggesting that there is a significantly (ANOVA  $F_{2,14} = 3.81$ ,  $p = 0.018$ ) higher potential for Fe reduction in this polygon type. Previous, lab-scale incubations using Barrow LC polygon soils showed that Fe reduction and methanogenesis occur concurrently<sup>12</sup> in the organic soils, supporting our findings. Our results suggest that Fe reduction capacity is not equally distributed across the polygonal landscape and, similar to methanogenesis, might be limited to the wetter LC polygons. By contrast, genes involved in the sulfur cycle were found across all polygon types and soil horizons in relatively similar abundances (Supplementary Fig. 21).

The relative abundances of genes involved in N cycling varied among polygon types and soil horizons. N-fixation genes were significantly higher in abundance ( $p < 0.01$ ) in LC polygons and mineral soils (Fig. 3c), whereas the HC and FC polygons had higher amounts (up to 0.1% of total gene content) of genes involved in nitrification; for the HC polygons this corresponds to the relatively high abundance of putative nitrifiers that were detected (Supplementary Fig. 4c). All polygons had low genomic potential for denitrification and N<sub>2</sub>O production. Genes for N-assimilation (*nirB*, *nrf*) from NO<sub>2</sub><sup>-</sup> to NH<sub>4</sub> were found in all polygons in similarly high abundance, as well as genes involved in conversion of organic N to NH<sub>4</sub> (for example, *gdh* and *ureC*). Our metagenome data are supported by gas flux measurements that found negligible amounts of N<sub>2</sub>O production (0.130–0.742 nmol N<sub>2</sub>O m<sup>-2</sup> s<sup>-1</sup>) at only 5 of 153 points across the polygon site at Barrow, with no apparent link to soil moisture.

**Novel microbial genomes across the polygons.** Thirty three nearly complete<sup>13</sup> (>89%) and low contamination<sup>13</sup> (<15%) draft microbial genomes were binned from the assembled metagenomic sequences (Fig. 4, Supplementary Table 5). The genome bin sizes were between 2.1 and 7.6 Mb with a variable GC content ranging from 36–69%. Most of the genomes constituted, on average, 1% (min. 0.25%, max. 5.96%) of the total reads in their respective metagenomes; in total 10–14% of the metagenome

reads could be mapped back to genome bins (Supplementary Table 5). We assembled full-length 16S sequences in half of the genome bins, ranging between 82 and 95% similarity to 16S sequences from previously sequenced genomes (Supplementary Table 6). *Proteobacteria*, *Bacteroidetes*, *Acidobacteria*, and

*Actinobacteria* were the four main phyla that the genome bins were assigned to. The majority of the low GC genomes were identified as *Bacteroidetes*, whereas most of the high GC genomes were represented in *Actinobacteria* (Supplementary Table 6, Supplementary Table 7).



Across different polygon types, genome bins were from genetically different but functionally similar microbial species. The bins showed low similarity to publicly available genomes (Fig. 4b) suggesting that they represent novel species. All of the genome bins contained a variety of genes encoding degradation of cellulose, hemicellulose, and polysaccharides. Almost all of the genomes contained glucosidases for degradation of oligosaccharides, but genes encoding the initial steps in cellulose breakdown were found in just some of the genomes (Fig. 4). The majority of the chitinases were located in *Bacteroidetes* and *Actinobacteria* genome bins. Hemicellulose and cellulose degradation potential were more common than chitinases across different phyla. The lower occurrence of complete cellulose breakdown pathways in Barrow genomes relative to those for utilization of cellobiose (Fig. 4) is consistent with previous research showing that cello-oligosaccharide utilization outnumbered cellulase-metabolic pathways in currently known genomes<sup>14</sup>.

Genomic analyses indicated that most of the genomes represented facultative anaerobes, with pathways for fermentation of glucose, lactate, and pyruvate (Figs. 4a and 5). Several of the genomes had multiple copies of alcohol and aldehyde dehydrogenases. Formate and CO-dehydrogenases were mainly detected in genomes binned from the HC and FC polygon metagenomes. Three genomes—one in each polygon type—also contained putative ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) genes for CO<sub>2</sub> assimilation. Since molecular hydrogen is an important metabolic intermediate in saturated environments (i.e., wetlands) we investigated the presence of hydrogenases in Barrow genomes. Fe,Fe-hydrogenase complexes, which are found in fermenters that produce high molar ratios of H<sub>2</sub><sup>15</sup>, were not detected in any of the genomes. Ni,Fe-hydrogenases that are commonly found in organisms using H<sub>2</sub> as a donor for respiratory metabolism<sup>16</sup> were mainly present in FC polygon genomes.

The genomes expanded upon our findings from the metagenomes regarding the potential for cycling of N. All of the genomes contained genes required for N-assimilation, and many also had genes encoding transporters for NH<sub>3</sub> and NO<sub>3</sub><sup>-</sup>/NO<sub>2</sub><sup>-</sup> (Fig. 5, Supplementary Table 7). Denitrification pathways were sparse and mostly incomplete across all polygon types. These functions were mainly found in FC polygons where four out of eight genomes contained incomplete denitrification pathways. In addition, one *Bacteroidetes* bin from a LC polygon contained an incomplete denitrification pathway while the remaining LC genomes did not comprise any known denitrification genes. Additionally, we did not detect genes for nitrification, N-fixation, or aerobic CH<sub>4</sub> oxidation in any of the genomes.

Data mining of the binned genomes also supported the metagenome data with respect to the potential for methanogenesis in the LC polygons. We assembled one partially complete methanogen (completeness 41.06%, contamination: 9.1%) from a LC polygon (Fig. 5). This new candidate Archaeal strain E.001 had 87% 16S similarity to a previously described

hydrogenotrophic methanogen, *Methanolinea tarda* strain NOBI-1<sup>17</sup>, and its *mcrA* gene had 76% identity (blastp e-value: 6E-58) to *Methanocella paludicola*, another hydrogenotrophic methanogen isolated from a rice paddy<sup>18</sup> (Rice Cluster I). Additionally, the genome bin contained a *mtd* (F420-dependent methylene tetrahydromethanoptertetrahydromethanopterin dehydrogenase; EC 1.5.99.9) gene which exhibits 56% identity (blastp e-value: 2E-104) to *Methanolinea tarda*<sup>17</sup>. This draft genome also contained the complete pathway for hydrogenotrophic methanogenesis (Fig. 5).

Additional genes identified in the genome bins included those encoding cold and heat shock proteins, alkaline phosphatases and sulfate adenylyltransferases (Figs. 4a and 5). A large number of cold-shock proteins have previously been identified in psychrophiles and cold-adapted microorganisms<sup>7</sup>. Therefore, phylum-wide availability of these genes was anticipated. Unlike the cold-shock response, heat shock can be triggered by nonspecific stress factors<sup>19</sup> which could further support the survival of arctic microorganisms under changing environmental conditions. Four out of eight *Actinobacteria* genome bins had genes for spore formation (Fig. 4a). Seven genomes had genes for flagella suggesting the potential for motility.

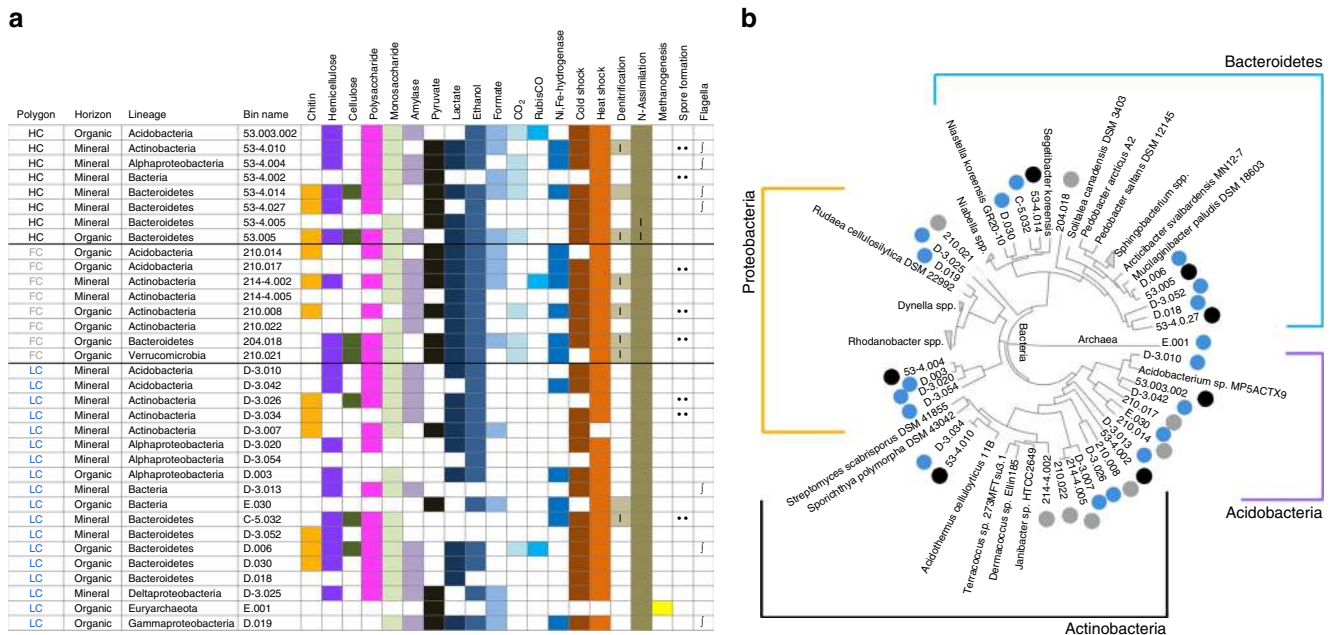
## Discussion

The majority of arctic microbiology research to date has focused on high soil moisture locations such as LC polygons<sup>20–24</sup>, as potential hot-spots for CH<sub>4</sub> production. A recent IPCC report<sup>4</sup> indicates a stronger climate forcing by CH<sub>4</sub> emissions (equivalent to 34 g of CO<sub>2</sub> per g of CH<sub>4</sub>) than previously anticipated. This comparison highlights the importance of CH<sub>4</sub> emissions from permanently inundated areas. However, as water is heterogeneously distributed spatially and temporally in the tundra, oxygen availability would also be expected to be unevenly distributed<sup>25</sup>. As a result aerobic and anaerobic processes alike are of importance to understanding GHG production from polygonal landscapes<sup>2</sup>. Disturbances in polygonal grounds, such as caused by thermo-erosion, can occur on a decadal scale and result in severe and immediate impacts on the terrain hydrology, ALT, and vegetation distribution<sup>26–28</sup>. Thus, understanding the distribution of the microbial and environmental controls over the polygonal landscape that regulates the emission of GHG is essential.

Here we demonstrated that microbial diversity and composition changed both horizontally across the polygonal landscape and with depth into the permafrost layer. A key feature of the permafrost community was an increasing dominance of *Actinobacteria* with depth, corresponding with a decline in carbon and nutrients (Fig. 2). In particular, *Solirubrobacterales* increased in abundance with depth (Supplementary Fig. 8a); presumably due to their ability to survive cold, nutrient limited conditions, such as the Dry Valleys of Antarctica<sup>7</sup>. *Actinomycetales* were otherwise the dominant order in the permafrost, and their co-occurrence with representatives of *Solirubrobacterales* may be due to niche partitioning<sup>29</sup>, as a survival strategy to utilize limited resources.

**Fig. 3** Metabolic potential predicted from metagenomes vary among polygon types. Clustering of this functional potential via **a** principal component analysis of the relative abundance of functional genes (FOAM-KO Level2) showed an association of metabolic pathways to each polygon type. The percentage variation explained by the principal components is indicated on the axes. Organic (closed) and mineral (open) soil depths were represented as circles, metabolic pathways are represented in diamond shape symbols. Two PCA axes could explain 41.9% of the observed variation. Labels highlight FOAM pathways that showed a strong correlation to the ordination of samples. **b** Relative abundance of CH<sub>4</sub> production (methyl coenzyme M reductase—*mcrABG*) and oxidation genes (particulate methane monooxygenase—*pmoABC*, soluble methane monooxygenase—*mmoXYZ* and methanol dehydrogenase—*mxoFJGD*), active layer thickness (ALT) was collected at the time of sampling for sequencing (09/24/2011); CO<sub>2</sub> and CH<sub>4</sub> fluxes were measured in two consecutive years, 2012 and 2013. Point locations for CO<sub>2</sub> and CH<sub>4</sub> flux measurements are represented in Fig. 1b, Supplementary Fig. 9 and Supplementary Fig. 10. In 2012, fluxes were measured from four locations per polygon ( $n = 4$ ) as a single time point measurement on 12 August 2012. Between July–October 2013, fluxes were measured monthly from center, rim, and trough of each polygon ( $n = 3$  per polygon)<sup>10</sup>. Error bars represent the standard error between measurements from same polygon. **c** Relative abundance of N cycle genes was also different among different polygons. Organic (closed) and mineral (open) soil depths were represented as circles sized to correspond to the relative abundance of each gene





**Fig. 4** Thirty-three nearly complete bacterial and one partial archaeal genome are binned from assembled soil metagenomic reads. **a** Distribution of the metabolic capacity for organic C utilization, fermentation, and respiration identified in the bacterial genomes reconstructed in this study. Dot sign shows presence of spore forming genes; integral sign denotes presence of flagellar movement genes; (I) designates incomplete pathways. **b** Maximum likelihood phylogenetic tree was constructed by using 49 highly conserved COG families from publicly available genomes

Our novel findings go beyond generalizations to illustrate which specific C decomposition pathways (e.g., genes for cellulose and chitin degradation; Fig. 4a) are present in the soil microbiome and the metabolic capacity of specific populations (genomes) from the site and how those data are correlated to gas flux measurements. In particular, there were contrasting carbon metabolic pathways in the HC and LC polygons, with a higher relative abundance of genes for degradation of more complex organic compounds in the HC polygons and for metabolism of simple carbohydrates and fermentation in the LC polygons. We also found significant differences in genes involved in the nitrogen cycle. For example, higher levels of genes for nitrogen fixation were detected in LC polygons and higher levels of genes for nitrification in HC polygons. Another novel finding was the link between low levels of genes for N<sub>2</sub>O production across the polygons that corresponded to low measured N<sub>2</sub>O flux at the Barrow site.

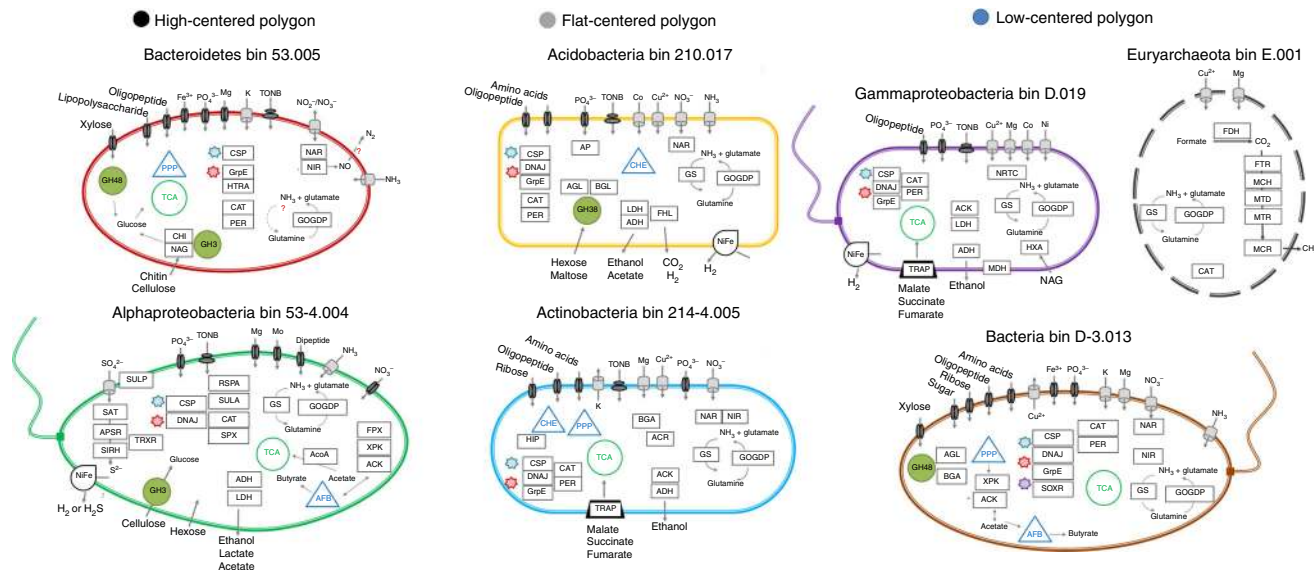
The balance between methane generation and oxidation was a key differentiator across the polygon transect. The key gene for methanogenesis was only observed in the LC polygons, corresponding to the significantly higher measured CH<sub>4</sub> flux at this location (Fig. 3b), as we previously reported for bog samples at a different site in Alaska<sup>30</sup>. Other genes involved in methanogenesis were found in all polygon types, but they had ~100× higher abundance in the LC polygons, compared to the HC and FC polygons (Supplementary Fig. 9). We hypothesize that methanogenesis is favored in the wetter regions of the LC polygons, not only because they retain water, but also because they have warmer seasonal temperatures<sup>1</sup> that delay freezing of top soils and result in accumulation of organic matter in the polygon centers. High amounts of organic material favor methanogenesis both by provision of substrates for fermentation and for maintaining low oxygen levels. CH<sub>4</sub> was previously measured in pore water samples collected below 20 cm depth from both HC and FC polygons at Barrow<sup>10</sup>. However, we found that the amount of CH<sub>4</sub> flux from wetter areas in HC and FC polygons was much lower than that measured from the LC polygons. We hypothesize

that the accumulated CH<sub>4</sub> was oxidized in the upper soil layers in the HC and FC polygons before it reached the surface. Our metagenome data suggest that methane oxidizers were more abundant in those polygon types, thus supporting this hypothesis. In the extreme case, in some mineral arctic soils, CH<sub>4</sub> flux can be entirely suppressed by CH<sub>4</sub> oxidation thus becoming a sink for atmospheric CH<sub>4</sub><sup>30</sup>.

One hypothesis is that hydrological conditions (i.e., lateral movement and drainage) and soil structure allow the presence of both micro-aerophilic and anaerobic conditions in the LC polygons, thus supporting both CH<sub>4</sub> production and oxidation. However, in the mineral soils of the LC polygons there was a significantly lower relative abundance of *pmo* genes compared to the rest of the samples. These samples also had the lowest abundance of *Methylobium* (Supplementary Fig. 2b) which was otherwise the most abundant CH<sub>4</sub> oxidizing bacteria in HC and FC polygons. The reduced potential for CH<sub>4</sub> oxidation (based on *pmo* gene abundance) could be one of the underlying mechanisms leading to high CH<sub>4</sub> fluxes<sup>10</sup> from this polygon type.

In inundated areas such as LC polygons 75–95% of the subsurface CH<sub>4</sub> has been reported to be lost to the atmosphere due to the combined upward transport of mediated by plants and ebullition<sup>31</sup>. During this transport only 0.8–8.7% of subsurface CH<sub>4</sub> is available for microbial oxidation<sup>31</sup>. The dominant vegetation type at our location was *Carex aquatilis*, a sedge with aerenchyma known to transport CH<sub>4</sub><sup>32</sup>. Thus despite the high potential for CH<sub>4</sub> production in the wetter LC polygons, vegetative transport and removal from the system may be factors that limit CH<sub>4</sub> oxidation by the resident microbes.

By targeting genes involved in cycling of N we provide the first metagenomic evidence that the biomes of arctic polygonal grounds are enriched in genes for N-assimilation. This finding suggests that the microbial communities are poised to take up N as it becomes available in this N poor environment<sup>33</sup>. Previously,  $\delta^{15}\text{N}$  measurements identified snow melt and nitrification potential in HC polygons as the main sources of NO<sub>3</sub><sup>-</sup> in Barrow polygons<sup>34</sup>. Although Lipson et al.<sup>21</sup> showed that denitrification



**Fig. 5** Barrow genome bins are indicative of metabolic flexibility and high potential for soil organic matter degradation. Metabolic models with detected genes (box) show a wide variety of functions in Barrow bins. Cell morphology is arbitrarily displayed. Pathways; PPP: pentose phosphate pathway, TCA: citric acid cycle, GG: glycolysis-glyconeogenesis, AFB: acetyl-CoA fermentation to butyrate, CHE: bacterial chemotaxis, TONB: ferric siderophore transport system. CAZymes; GH3: glycoside hydrolase family 3, GH38: glycoside hydrolase family 38, GH48: glycoside hydrolase family 48. Methanogenesis; FDH: formate dehydrogenase (EC 1.2.1.2), FTR: formylmethanofuran--tetrahydromethanopterin N-formyltransferase (EC: 2.3.1.101), MCH: methenyltetrahydromethanopterin cyclohydrolase (EC: 3.5.4.27), MTD: methylenetetrahydromethanopterin dehydrogenase (EC:1.5.98.1), MCR: methyl-coenzyme M reductase (EC:2.8.4.1), MTR: tetrahydromethanopterin S-methyltransferase (EC:2.1.1.86). Stress Response; CSP: cold shock protein, DNAJ: heat shock protein/chaperone protein, GrpE: heat shock protein, SOXR: redox-sensitive transcriptional activator SoxR, RSPA: starvation sensing, SULA: cell division inhibitor, SPX: superoxide dismutase [Fe] (EC 1.15.1.1), HIP: HipA protein, inhibits cell growth and induces persistence, ACR: arsenic resistance protein. Assimilatory sulfate reduction. Genes; SULP: sulfate permease, SAT: sulfate adenyllyltransferase (EC 2.7.7.4), APSR: adenyllyl-sulfate reductase [thioredoxin] (EC 1.8.4.10), SIRH: sulfite reductase [NADPH] (EC 1.8.1.2), TRXR: thioredoxin reductase (EC 1.8.1.9), PER: peroxidase (EC 1.11.1.7), CAT: catalase (EC 1.11.1.6), HTRA: protease, RUB: rubrerythrin, CYC-c551: cytochrome c551 peroxidase (EC 1.11.1.5), FFX: fructose-6-phosphate phosphoketolase (EC 4.1.2.22), XPK: xylulose-5-phosphate phosphoketolase (EC 4.1.2.9), ACK: acetate kinase (EC 2.7.2.1), ACS: acetyl-CoA synthetase (EC 6.2.1.1), NAR: assimilatory nitrate reductase (EC 1.7.99.4), NIR: ferredoxin nitrite reductase (EC 1.7.7.1), GS: glutamine synthetase (EC 6.3.1.2), GOGDP: glutamate synthetase (EC 1.4.1.13), AGL: alpha-glucosidase (EC 3.2.1.20), BGL: beta-glucosidase (EC 3.2.1.21), BGA: beta-galactosidase (EC 3.2.1.23), AXYL: alpha-xylosidase (EC 3.2.1.-), BXYL: beta-xylosidase (EC 3.2.1.37), CHI: chitinase (EC 3.2.1.14), NAG: beta-N-acetylglucosaminidase (EC: 3.2.1.96), ADH: alcohol dehydrogenase (EC 1.1.1.1), LDH: lactate dehydrogenase (EC 1.1.1.28), MDH: methanol dehydrogenase (EC:1.1.2.7), AP: alkaline phosphatase (EC 3.1.3.1), FHL: formate hydrogenlyase (EC 1.2.1.2), NRTC: nitrogen regulatory protein C (NtrC), HXA: hexosaminidase (EC 3.2.1.52)

potential was present at the Barrow site; our metagenome findings suggest that  $\text{NO}_3^-$  was utilized as a N source, but not lost through denitrification. Also we observed negligible amounts of  $\text{N}_2\text{O}$  flux from this landscape with no apparent links to polygon types, soil moisture or other environmental conditions (Fig. 3b).

The metagenome data suggest that the wetter LC polygons are favorable for both methanogenesis and Fe reduction processes (Supplementary Fig. 20). Previously, lab-scale incubation studies from Barrow LC polygons showed that Fe reduction and methanogenesis occur concurrently<sup>12</sup> in the organic soils, supporting our findings. Microbes involved in both processes are potential consumers of available organic acids and fermentation products<sup>35</sup> and we found that the LC polygons were enriched with genes involved in degradation of simpler carbohydrates (i.e., sugars) and mixed acid fermentation pathways (Supplementary Fig. 15, Supplementary Fig. 16) that could be linked to these processes. A combination of anaerobic conditions and availability of organic acids from root exudates from sedges<sup>36,37</sup> are likely facilitators of microbial potential for fermentation in the wet areas.

Increased  $\text{CH}_4$  emissions with permafrost thaw is a common observation in arctic tundra<sup>38-40</sup>. However, the influence of increasing soil temperatures on GHGs in polygonal grounds depends strongly on the interactions between temperature, soil moisture and drainage potential<sup>10</sup>. At BEO,  $\text{CH}_4$  flux was predominantly observed from wetter, LC polygons (Fig. 3b). However, despite the differences in soil moisture distribution

(Supplementary Table 8), we did not observe significant differences among  $\text{CO}_2$  fluxes in FC and LC polygons; although the HC polygons had intermittently higher fluxes. It should also be noted that there was a seasonal difference in  $\text{CO}_2$  fluxes, with highest fluxes in the late summer months (Supplementary Fig. 9, Supplementary Fig. 10)<sup>10</sup>. These findings could be due to differences in soil organic matter deposition, decomposition and root respiration rates over the season. In the HC polygons there were several genes encoding enzymes for degradation of C polymers that were significantly (~10x) more abundant when compared to the other polygon types. These included genes encoding: xylan 1,4-beta-xylosidase (*xynB*) and chitinase (*chiA*) (Supplementary Fig. 11). This enrichment of hydrolytic enzymes suggests that complex plant polymers are available as microbial growth substrates in HC polygons. In contrast, the wetter LC polygons were enriched with genes for anaerobic processes, such as sugar and mixed acid fermentation, iron reduction, and methanogenesis (Supplementary Fig. 11, Supplementary Fig. 16, Supplementary Fig. 18, Supplementary Fig. 20), some of which resulted in anaerobic  $\text{CO}_2$  fluxes. As polygonal landscapes transition into a drier and more high-centered state<sup>26,41,42</sup>, we hypothesize that the corresponding decrease in soil moisture that leads to death of vascular plants will at least transitionally provide plant residues that serve as a substrate for the resident soil microbes.

We binned several novel draft genomes from the metagenomes (Figs. 4 and 5). These bins were both taxonomically and

functionally diverse and contained multiple genes involved in breakdown and utilization of diverse structural polysaccharides (Figs. 4 and 5). This wide range of substrate utilization capacity has been observed in several microbial isolates from the Arctic<sup>43</sup> and hypothesized to be related to improved resilience to fluctuating temperatures and nutrient-deficient conditions in tundra soils. Our results, moreover, point towards functional heterogeneity as a mechanism to cope with changing moisture conditions because facultative metabolism was found across the genome bins. This potential was accompanied by presence of cellulose debranching enzymes and alcohol dehydrogenases both in organic and mineral soil layers (Figs. 4a and 5). In arctic peat soils gradients of aerobic to anaerobic conditions have been hypothesized to structure microbial metabolic potential<sup>21,36</sup>. Based on our findings we also hypothesize that soil structure and drainage create dry or wet micro pockets within the same soil horizon that favor the growth and survival of taxonomically diverse microorganisms with facultative metabolism.

We assembled one partially complete Archaeal bin that based on 16S and *mcrA* gene identification showed low similarity to previously described hydrogenotrophic methanogens<sup>18,37</sup>. This methanogen was co-localized with fermentative bacteria (Fig. 4a) that can produce H<sub>2</sub> thus potentially supporting hydrogenotrophic methanogenesis. However, the metagenome results showed that multiple Archaeal species and pathways were potentially involved in production of CH<sub>4</sub> within LC polygons (Supplementary Fig. 11). In fact, in seasonal succession of methanogenesis, the acetoclastic pathway was found to be dominant when active layer soils were thawed<sup>35</sup> and hydrogenotrophic methanogenesis to be dominant at the end of summer<sup>31</sup>. This methanogen represents one of many Archaeal genomes that were present in these samples.

We also screened the genome bins for genes involved in stress survival. Spore forming was not widely detected in Barrow genome bins (Fig. 4a); supporting the previous assessment<sup>44</sup> that spores are not the best survival strategy for freezing conditions. We hypothesize that instead of dormancy and spore formation that the microorganisms at Barrow use a combination of cold and heat shock proteins, cryoprotectants (including sugars originating from soil organic matter<sup>45</sup>) and low activity—enabling critical cellular functions such as DNA repair—to survive harsh winter conditions.

In summary, the structure and functions of soil microbial communities in polygonal grounds are ultimately shaped by the landscape topography. Our research demonstrates that a metagenomic approach can resolve metabolic roles for uncultivated soil microorganisms and provide the framework needed to predict their metabolic response when the landscape evolves. Our results demonstrate that distribution of genes involved in soil organic C degradation is significantly related to the landscape topography. This knowledge is important because it provides evidence towards predicting the microbial metabolic response to increased temperatures in arctic soils that can potentially contribute to a positive climate feedback<sup>39,46,47</sup>. Our results advances previous findings of genes and pathways involved in soil organic C degradation in other arctic soil metagenomes<sup>21,37,48–51</sup>. Specifically, the knowledge gained provides more mechanistic detail about the metabolic potential of the soil microbiome and can help to define the metabolic routes for GHG production for refinement of model predictions. The combination of detailed hydrological, biogeochemical, vegetation, and microbial data obtained from the BEO site<sup>2,10,52,53</sup> suggest that under a warmer climate, microbial contributions to GHG emissions from these ecosystems will ultimately depend on how soil moisture (i.e., wetter vs drier<sup>6</sup>) is distributed across the landscape.

## Methods

**Sampling location and characteristics.** Site selection, sample collection, and sample characterization were previously reported<sup>1</sup>. Soil and permafrost samples were collected from the BEO which is located approximately 6 km east of Barrow, AK (71.3° N, 156.5° W). This Arctic Coastal Plain research site contains thaw lakes, drained thaw lake basins, and interstitial ice-wedge polygons that cover more than half of the land surface<sup>41</sup>. Active layer soils in the BEO are classified as Gelisols and characterized by an organic-rich surface layer underlain by a horizon of mineral soil and a frozen organic-rich mineral layer. In total 31 soil samples from identified organic and mineral layers of HC, FC, and LC polygons were collected prior to seasonal freeze on 24 September 2011. During the sampling period, peak plant growth and active layer thaw depths were also measured. The LC polygons at this location are characterized by narrow, saturated troughs, relatively high and dry topographic ridges, and depressed year-round saturated centers. By comparison, the HC polygons have seasonally saturated troughs, dry ridges and dome shaped high and dry centers. FC polygons contain seasonally saturated troughs that border the polygon with elevated but flat centers. Mosses and liverworts are widely found in all polygon types. *Carex aquatilis* (sedge), *Eriophorum* (sedge), *Petasites frigidus* (perennial herb), *Luzula nivalis* (perennial herb), and *Dupontia fisheri* (sedge) are among most commonly observed plants at this location<sup>53</sup>. Among sedges *Carex aquatilis* was shown to be more abundant in LC polygons whereas *Eriophorum* and *Dupontia fisheri* were more abundant in FC polygons<sup>10</sup>. The mean canopy height across different polygon features was lower in HC polygons (1.9–10.8 cm) in comparison to FC (4.1–11.3 cm) and LC (7.9–15.5 cm) polygons.

Upon collection, the samples were kept frozen at –20 °C. Thaw depth measurements were obtained using a steel probe with centimeter gradations starting at the bottom of the probe. Since the measurements were collected at the end of the growing season, the thaw depth measurements are considered to be equivalent to the ALT. Average active layer depth along this transect is 0.37 m. Samples were analyzed to determine density, C and N content, pH, electrical conductivity (μS), and moisture. Additionally, major and trace cations, including Ca<sup>2+</sup>, Na<sup>+</sup>, Mg<sup>2+</sup>, Fe<sup>3+</sup>, Al<sup>3+</sup>, Si<sup>4+</sup>, K<sup>+</sup>, Mn<sup>2+</sup>, Cu<sup>2+</sup>, and Zn<sup>2+</sup> were measured from extracted pore water<sup>1</sup>. During sapling, soil temperatures were between 1.2 and 1.8 °C at all sampling locations. Active layer (0.0–0.53 m depth) soil temperatures fluctuate between –22 °C and 5 °C over the course of a year<sup>52</sup>. HC polygons were shown to contain a higher amount of C (23.2 ± 5.4%) relative to other polygons (FC: 12.9 ± 5.1% and LC: 15.4 ± 8.4%)<sup>2</sup>. FC polygons were cooler than the other polygons<sup>2</sup>. LC polygons had the deepest average ALT (HC: 35 ± 3 cm, FC: 34 ± 5 cm, and LC: 45 ± 7 cm) and the highest Fe<sup>3+</sup> concentrations (HC: 3.1 ± 1.9 μg l<sup>-1</sup>, FC: 3.1 ± 1.9 μg l<sup>-1</sup> and LC: 3.4 ± 2.9 μg l<sup>-1</sup>)<sup>2</sup>. The soil samples were freeze-dried over a 7 day period until no further weight loss was observed. The gravimetric moisture content was calculated from the weight difference before and after the freeze-drying process. Note that soils in the BEO are highly porous and have a high water holding capacity<sup>1</sup>. The average gravimetric soil moisture (g water gr soil<sup>-1</sup>) was highest in the LC polygons; values for all sampling locations were: HC Organic: 0.71 ± 0.10, Mineral: 0.59 ± 0.05; FC Organic: 0.76 ± 0.07, Mineral: 0.41 ± 0.12; LC Organic: 0.85 ± 0.05, Mineral: 0.50 ± 0.05<sup>1</sup> (Supplementary Table 8). When observed, water levels in troughs were measured from the water surface to the top of the active layer. Dissolved oxygen concentrations measured in nearby locations showed 86–92% O<sub>2</sub> in surface and at 31–46% O<sub>2</sub> in mineral soil depths<sup>31</sup>.

Permafrost at the site is continuous, ice-rich, and is present to depths greater than 350 m<sup>54</sup>. Duplicate permafrost cores were obtained via drilling with a 3 inch (7.62 cm) diameter SIPRE corer in the center of a FC polygon (154 m from the south end of the transect) in April of 2012. The total lengths of the cores are 1.00 and 2.65 m. This is the deepest and most intact core pair that was obtained during this sampling campaign. The permafrost temperature range at this location is between –17 °C and –1 °C at 0.5 m to –12 °C and –5 °C at 2.65 m<sup>52</sup>. Cores were stored frozen at –20 °C and scanned using a modified Seimens Somatom HiQ third-generation medical computed tomography (CT) scanner<sup>55</sup>. X-ray CT scanning is a nondestructive method that provides the three-dimensional density (g cm<sup>-3</sup>) distribution of the scanned samples in order to differentiate ice content from higher density materials such as organics or clays<sup>52</sup>. After scanning, cores were subsectioned while frozen under sterile conditions with a one inch (2.54 cm) diameter mini-corer in 5 cm depth intervals. Each depth interval was divided into four sections, homogenized and analyzed independently. The rest of the core was subsequently sliced with a chop saw into disks for chemical analysis (water chemistry, C and N content, Supplementary Table 1). Anions and cations were measured in pore water sampled with rhizon samplers and analyzed with IC and ICP-MS system as described previously<sup>1</sup>.

**DNA extraction and 16S analysis.** Total DNA was extracted from 2 g material as described earlier<sup>56</sup>. Extractions were purified using an AllPrep DNA/RNA kit (Qiagen, Valencia, CA, USA) and were quantified using Quant-iT dsDNA HS assay kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's manual. PCR reactions were performed as described earlier<sup>57</sup>. Active layer samples were sequenced on Illumina GA-Ix platform (at the Lawrence Berkeley National Laboratory, CA, USA) resulting in ~95 bp single-end reads and permafrost samples were sequenced on Illumina MiSeq platform (at JGI, CA, USA), resulting in ~250 bp paired end reads.

The paired-end sequences were overlapped and merged using FLASH<sup>58</sup>. Quality filtering and demultiplexing were performed as described previously<sup>59</sup>. Sequences were grouped into operational taxonomic units (OTUs) based on 97% sequence identity, and chimeric sequences were removed, using UPARSE<sup>60</sup>. OTUs were given taxonomic assignments in QIIME<sup>61</sup> version 1.7.0 using RDP classifier<sup>62</sup> and Greengenes database release 13\_5. Phylogenetic trees were created using FastTree<sup>63</sup> under QIIME's default parameters and these trees were used for the calculation of alpha- (Shannon's H', Chao1 and Faith's PD) and beta-diversity (weighted UniFrac distance) metrics. Variable sequencing depth was normalized across samples, by performing a single rarefaction at a depth of 7500 sequences per active layer soil or 22,400 sequences per permafrost core. The weighted UniFrac distance matrices were used to visualize the community composition.

**Metagenome sequencing and analysis.** Sixteen metagenomic shotgun sequencing libraries were prepared to represent three polygon features (rim, trough, and center) and two soil horizons (organic and mineral). We generated metagenomes only from active layer soil samples. Extracted total genomic DNA (300 ng) was used in sequencing library preparation as described earlier<sup>57</sup>. Three samples were multiplexed per lane of Illumina HiSeq 2000 (Yale Center for Genome Analysis, CT). Paired-end raw reads were cleaned for sequencing adaptor and filtered for low quality reads with CLC Workbench 5.1 resulting (minimum quality cutoff of 3) up to 30 Gb of high-quality sequence data of 92 bp in length (Supplementary Table 2).

Prodigal<sup>64</sup> was used to predict coding regions from the single reads. The translated proteins from all detected coding regions of each metagenome were annotated by searching against the FOAM (Functional Ontology Assignments for Metagenomes)<sup>65</sup> database of hidden Markov models generated from Pfam profiles. This screening resulted in the annotation of an average 2.5E+06 to 1.0E+07 genes per sample. Gene abundances (gene counts per KEGG Orthology—KO) were normalized to the number of recombinase A copies detected in each metagenome. In addition, each sample was individually assembled with MEGAHIT<sup>66</sup>. Genome fragments that were larger than 1 kb were clustered using MaxBin<sup>67</sup>. Potential mis-binnings were identified with CheckM<sup>13</sup> where genome bin completeness and contamination are reported (Supplementary Table 4). Protein-coding genes associated with a genome bin were manually checked to have the same phylogenetic affiliation, guanine-cytosine content and to contain 16S ribosomal RNA (rRNA) gene—where available—or phylogenetic marker genes. AMPHORA2<sup>68</sup> was used to evaluate taxonomy of bins assigned to Bacteria with Maxbin and CheckM. Marker lineage was reported if 75% of the classifications were in agreement at a particular taxonomic level. Genomes were annotated using Department of Energy Systems Biology Knowledgebase (KBase - <http://kbase.us>) and RAST<sup>69</sup> servers. Binned genomes were compared to publicly available genomes as implemented in KBase where a set of reference alignments based on 49 highly conserved COG families is used to find the matching corresponding set of sequences in the binned genomes. The sequences from the selected genome were then inserted into the reference alignments and the tree was rendered from them using FastTree<sup>63</sup>.

**In situ GHG flux measurements.** Details of flux measurements are described in Vaughn et al. (2016)<sup>10</sup>. Briefly, the net CO<sub>2</sub> and CH<sub>4</sub> fluxes from polygons were measured with an opaque chamber (25 cm diameter), connected to a Los Gatos Research, Inc. (LGR) portable Greenhouse Gas Analyzer, placed on a PVC base (installed approximately 15 cm deep) for 4–8 min. In inundated plots, a floating chamber whose base extended 4 cm below the water surface was used. Fluxes were calculated from the slope of the linear section of the LGR plot of GHG concentration versus time. CO<sub>2</sub> and CH<sub>4</sub> fluxes were collected in two consecutive years, 2012 and 2013. Point locations for CO<sub>2</sub> and CH<sub>4</sub> flux measurements are represented in Fig. 1b. In 2012, fluxes were measured from four locations per polygon ( $n = 4$ ); three replicate measurements from the polygon centers and one measurement from polygon troughs in HC and FC polygons and a polygon rim in LC polygons as a single time point measurement on 12 August 2012. In 2013, 1 × 1 m plots were established in center, rim, and trough of each polygon ( $n = 3$  per polygon)<sup>10</sup>. Fluxes of CO<sub>2</sub> and CH<sub>4</sub> were measured on 10–12 July, 7–16 August, 4–7 September, and 2–4 October 2013<sup>10</sup>. N<sub>2</sub>O fluxes were measured using a static opaque chamber (25 cm diameter) placed on a similar PVC base. N<sub>2</sub>O fluxes were collected from adjacent polygons to the ones studied here. In 1–3 July, 5–11 August, and 7–12 September 2012 chamber measurements were collected from center, rim, and trough ( $n = 3$  per polygon) of 16 polygons. Samples of headspace gas were collected via syringe through septa on the chamber at 10 min intervals over 40 min, and analyzed within 24 h using a Shimadzu GC-2014 electron capture detector.

**Statistical analysis.** All statistical tests were produced by using ade4 and vegan packages in the R statistical environment (<http://www.r-project.org>). Results are defined to be significant at  $p < 0.05$ . Ordination of the whole community detected by 16S amplicon sequencing was created from UniFrac matrix calculated by QIIME<sup>61</sup> software and presented in a principal coordinates analysis plot. Analysis of Similarity (ANOSIM) was used to test differences in community structure within the samples (1000 Monte Carlo permutations). Permutational multivariate analysis of variance (Adonis) tests were performed on weighted UniFrac values to determine the amount of variation explained by soil chemistry. Pearson ( $r$ ) correlation

and Spearman ( $\rho$ ) coefficients were determined between microbial abundances, geographical factors, and soil chemistry to test the alternative hypothesis that relative abundance of each prokaryotic phylum or gene is positively associated with geographical factors and soil chemistry. Moran's I; which calculates the degree of correlation between observations as a function of the spatial distance separating them, was used as implemented in ncf package to determine the presence of spatial autocorrelation in the distribution of microbial communities across different polygons. Microbial network analyses were performed in R using the phyloseq<sup>70</sup>. The network is based on the Bray-Curtis distance measure, and only connected nodes with a distance  $< 0.4$  were shown. In metagenomes, to test differences between relative abundances of different genes amongst polygons one-way ANOVA followed by the Tukey's honestly significant difference test was used. The differences among CO<sub>2</sub> and among CH<sub>4</sub> flux measurements were tested based on Akaike information criterion (AIC) values and Type III Sums of Squares using lme4 package to accommodate unbalanced data structure.

**Data availability.** The authors declare that the data supporting the findings of this study are available in this article and its Supplementary Information Files, or from the corresponding authors on request. All sequences have been deposited to the European Nucleotide Archive PRJEB20765, besides the MiSeq sequence data from permafrost samples which is available from the JGI Genome Portal (Project ID: 1019880).

Received: 24 May 2017 Accepted: 18 January 2018

Published online: 22 February 2018

## References

- Hubbard, S. S. et al. Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets. *Hydrogeol. J.* **21**, 149–169 (2013).
- Wainwright, H. M. et al. Identifying multiscale zonation and assessing the relative importance of polygon geomorphology on carbon fluxes in an Arctic Tundra Ecosystem. *J. Geophys. Res.: Biogeosci.* **120**, 788–808 (2015).
- Hinkel, K. M. et al. Spatial extent, age, and carbon stocks in drained taw lake basins on the Barrow Peninsula, Alaska. *Arct. Antarct. Alp. Res.* **35**, 291–300 (2003).
- Kirtman, B. et al. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to AR5.* 953–1028 (Cambridge University Press, Cambridge, UK and New York, NY, 2013).
- Romanovsky, V. E., Smith, S. L. & Christiansen, H. H. Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: a synthesis. *Permafrost. Periglac. Process.* **21**, 106–116 (2010).
- Schuur, E. A. G. et al. Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015).
- Jansson, J. K. & Taş, N. The microbial ecology of permafrost. *Nat. Rev. Microbiol.* **12**, 414–425 (2014).
- Sitch, S. et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653–679 (2015).
- Brioukhanov, A. & Netrusov, A. Aerotolerance of strictly anaerobic microorganisms and factors of defense against oxidative stress: a review. *Appl. Biochem. Microbiol.* **43**, 567–582 (2007).
- Vaughn, L. J., Conrad, M. E., Bill, M. & Torn, M. S. Isotopic insights into methane production, oxidation, and emissions in Arctic polygon tundra. *Glob. Change Biol.* **10**, 3487–3502 (2016).
- Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
- Roy Chowdhury, T. et al. Stoichiometry and temperature sensitivity of methanogenesis and CO<sub>2</sub> production from saturated polygonal tundra in Barrow, Alaska. *Glob. Change Biol.* **21**, 722–737 (2014).
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- Berlemont, R. & Martiny, A. C. Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* **79**, 1545–1554 (2013).
- Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu. Rev. Microbiol.* **66**, 429–452 (2012).
- Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
- Imachi, H. et al. Methanolinea tarda gen. nov., sp. nov., a methane-producing archaeon isolated from a methanogenic digester sludge. *Int. J. Syst. Evol. Microbiol.* **58**, 294–301 (2008).
- 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).
19. Casanueva, A., Tuffin, M., Cary, C. & Cowan, D. A. Molecular adaptations to psychrophily: the impact of 'omic' technologies. *Trends Microbiol.* **18**, 374–381 (2010).
  20. Frank-Fahle, B. A., Yergeau, É., Greer, C. W., Lantuit, H. & Wagner, D. Microbial functional potential and community composition in permafrost-affected soils of the NW Canadian Arctic. *PLoS ONE* **9**, e84761 (2014).
  21. Lipson, D. A. et al. Metagenomic Insights into Anaerobic Metabolism along an Arctic Peat Soil Profile. *PLoS ONE* **8**, e64659 (2013).
  22. Barbier, B. A. et al. Methane-cycling communities in a permafrost-affected soil on Herschel Island, Western Canadian Arctic: active layer profiling of *mcrA* and *pmoA* genes. *Fems. Microbiol. Ecol.* **82**, 287–302 (2012).
  23. Liebner, S., Harder, J. & Wagner, D. Bacterial diversity and community structure in polygonal tundra soils from Samoylov Island, Lena Delta, Siberia. *Int. Microbiol.* **11**, 195–202 (2008).
  24. Lipson, D. A. et al. Changes in microbial communities along redox gradients in polygonized Arctic wet tundra soils. *Environ. Microbiol. Rep.* **7**, 649–657 (2015).
  25. 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).
  26. Fortier, D., Allard, M. & Shur, Y. Observation of rapid drainage system development by thermal erosion of ice wedges on Bylot Island, Canadian Arctic Archipelago. *Permafrost. Periglac. Process.* **18**, 229–243 (2007).
  27. Liljedahl, A. et al. Nonlinear controls on evapotranspiration in arctic coastal wetlands. *Biogeosciences* **8**, 3375 (2011).
  28. Liljedahl, A. K. et al. Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology. *Nat. Geosci.* **9**, 312–318 (2016).
  29. Williams, R. J., Howe, A. & Hofmocker, K. S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* **5**, 358 (2014).
  30. Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **521**, 208–212 (2015).
  31. Lau, M. C. et al. An active atmospheric methane sink in high Arctic mineral cryosols. *ISME J.* **9**, 1880–1891 (2015).
  32. Throckmorton, H. M. et al. Pathways and transformations of dissolved methane and dissolved inorganic carbon in Arctic tundra watersheds: evidence from analysis of stable isotopes. *Glob. Biogeochem. Cycles* **29**, 1893–1910 (2015).
  33. Popp, T. J., Chanton, J. P., Whiting, G. J. & Grant, N. Methane stable isotope distribution at a Carex dominated fen in north central Alberta. *Glob. Biogeochem. Cycles* **13**, 1063–1077 (1999).
  34. Sistla, S. A., Asao, S. & Schimel, J. P. Detecting microbial N-limitation in tussock tundra soil: implications for Arctic soil organic carbon cycling. *Soil. Biol. Biochem.* **55**, 78–84 (2012).
  35. Heikoop, J. M. et al. Isotopic identification of soil and permafrost nitrate sources in an Arctic tundra ecosystem. *J. Geophys. Res.: Biogeosci.* **120**, 1000–1017 (2015).
  36. Rawat, S. R., Männistö, M. K., Bromberg, Y. & Häggblom, M. M. Comparative genomic and physiological analysis provides insights into the role of Acidobacteria in organic carbon utilization in Arctic tundra soils. *Fems. Microbiol. Ecol.* **82**, 341–355 (2012).
  37. Tveit, A. T., Urich, T., Frenzel, P. & Svenning, M. M. Metabolic and trophic interactions modulate methane production by Arctic peat microbiota in response to warming. *Proc. Natl Acad. Sci. USA* **112**, E2507–E2516 (2015).
  38. Herndon, E. M. et al. Pathways of anaerobic organic matter decomposition in tundra soils from Barrow, Alaska. *J. Geophys. Res.: Biogeosci.* **120**, 2345–2359 (2015).
  39. Johnston, C. E. et al. Effect of permafrost thaw on CO<sub>2</sub> and CH<sub>4</sub> exchange in a western Alaska peatland chronosequence. *Environ. Res. Lett.* **9**, 085004 (2014).
  40. Natali, S. M. et al. Permafrost thaw and soil moisture driving CO<sub>2</sub> and CH<sub>4</sub> release from upland tundra. *J. Geophys. Res.: Biogeosci.* **120**, 525–537 (2015).
  41. McCalley, C. K. et al. Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478–481 (2014).
  42. Jorgenson, M. T., Shur, Y. L. & Pullman, E. R. Abrupt increase in permafrost degradation in Arctic Alaska. *J. Geophys. Res.: Biogeosci.* **33**, L02503 (2006).
  43. Godin, E., Fortier, D. & Coulombe, S. Effects of thermo-erosion gully on hydrologic flow networks, discharge and soil loss. *Environ. Res. Lett.* **9**, 105010 (2014).
  44. Mondav, R. et al. Discovery of a novel methanogen prevalent in thawing permafrost. *Nat. Commun.* **5**, 3212 (2014).
  45. Vorobyova, E. et al. The deep cold biosphere: facts and hypothesis. *Fems. Microbiol. Rev.* **20**, 277–290 (1997).
  46. Taş, N. et al. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J.* **8**, 1904–1919 (2014).
  47. Xue, K. et al. Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. *Nat. Clim. Change* **6**, 595–600 (2016).
  48. Lawrence, D., Koven, C., Swenson, S., Riley, W. & Slater, A. Permafrost thaw and resulting soil moisture changes regulate projected high-latitude CO<sub>2</sub> and CH<sub>4</sub> emissions. *Environ. Res. Lett.* **10**, 094011 (2015).
  49. Yergeau, E., Hogue, 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).
  50. Taş, N. et al. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J.* **8**, 1904–1919 (2014).
  51. Mackelprang, R. et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**, 368–371 (2011).
  52. Dafflon, B. et al. Geophysical estimation of shallow permafrost distribution and properties in an ice-wedge polygon-dominated Arctic tundra region. *Geophysics* **81**, WA247–WA263 (2016).
  53. Sloan, V. et al. *Plant community composition and vegetation height, Barrow, Alaska, ver. 1*. Next Generation Ecosystem Experiments Arctic Data Collection; Carbon Dioxide Information Analysis Center; Oak Ridge National Laboratory: Oak Ridge, TN, USA, <https://doi.org/10.5440/1129476> (2014).
  54. Sellmann, P. V., Brown, J., Lewellen, R. I., McKim, H. & Merry, C. *The Classification and Geomorphic Implications of Thaw Lakes on the Arctic Coastal Plain, Alaska* Research Report 344 (Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, 1975).
  55. Kneafsey, T. J. & Moridis, G. J. X-Ray computed tomography examination and comparison of gas hydrate dissociation in NGHP-01 expedition (India) and Mount Elbert (Alaska) sediment cores: experimental observations and numerical modeling. *Mar. Petrol. Geol.* **58**, 526–539 (2014).
  56. DeAngelis, K. M., Lindow, S. E. & Firestone, M. K. Bacterial quorum sensing and nitrogen cycling in rhizosphere soil. *Fems. Microbiol. Ecol.* **66**, 197–207 (2008).
  57. Taş, N. et al. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J.* **8**, 1904–1919 (2014).
  58. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
  59. Bokulich, N. A. et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10**, 57–59 (2012).
  60. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
  61. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
  62. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
  63. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
  64. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
  65. Prestat, E. et al. FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res.* **42**, e145 (2014).
  66. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
  67. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 1–18 (2014).
  68. Kerepesi, C., Bánky, D. & Grolmusz, V. AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene* **533**, 538–540 (2014).
  69. Overbeek, R. et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2014).
  70. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).

## Acknowledgements

The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the DOE Office of Science. The Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported through contract number DE-AC0205CH11231 to Lawrence Berkeley National Laboratory. We acknowledge Stan Wullschlegel at Oak Ridge National Laboratory as the NGEE Project PI. Additional support was provided by JGI under CSP 2013 1044 “Next Generation Ecosystem Experiment (NGEE) in the Arctic” and the Microbiomes in Transition Initiative LDRD Program at the Pacific Northwest National Laboratory, a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC06-76RL01830.

### Author contributions

N.T., S.S.H. and J.K.J conceived the study and designed the experiments; N.T. prepared samples for 16S rRNA gene and metagenome sequencing, N.T. and E.P. performed data analysis, N.T., S.W. and Y.W. sterile sub-sectioned permafrost samples and performed DNA extraction, C.U. and S.S.H. collected active layer soil and permafrost core samples, T.K. provided CT-scanning and analysis of the permafrost core, S.G.T. facilitated 16S rRNA gene sequencing of permafrost samples, M.S.T. provided in situ gas-flux measurements, N.T. and J.K.J wrote the manuscript with feedback from all authors.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-03089-z>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018