FEMS Microbiology Ecology, 94, 2018, fiy110

doi: 10.1093/femsec/fiy110
Advance Access Publication Date: 14 June 2018
Research Article

RESEARCH ARTICLE

Dispersal limitation and thermodynamic constraints govern spatial structure of permafrost microbial communities

Eric M. Bottos[1,2,*], David W. Kennedy[1], Elvira B. Romero[1], Sarah J. Fansler[1], Joseph M. Brown[3], Lisa M. Bramer[4], Rosalie K. Chu[5], Malak M. Tfaily[5], Janet K. Jansson[1] and James C. Stegen[1]

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, 99352, USA, [2]Department of Biological Sciences, Thompson Rivers University, 805 TRU Way, Kamloops, BC, V2C 0C8, Canada, [3]Computational Biology, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, 99352, USA, [4]National Security Directorate, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, 99352, USA and [5]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, 99352, USA

*Corresponding author: Thompson Rivers University, 805 TRU Way, Kamloops, BC, V2C 0C8, Canada. E-mail: ebottos@tru.ca

One sentence summary: This work demonstrates that microbial distributions in permafrost exhibit spatial distributions that reflect dispersal limitation and selective pressures arising from thermodynamic constraints of the permafrost environment.

Editor: Max Haggblom

ABSTRACT

Understanding drivers of permafrost microbial community composition is critical for understanding permafrost microbiology and predicting ecosystem responses to thaw. We hypothesize that permafrost communities are shaped by physical constraints imposed by prolonged freezing, and exhibit spatial distributions that reflect dispersal limitation and selective pressures associated with these physical constraints. To test this, we characterized patterns of environmental variation and microbial community composition in permafrost across an Alaskan boreal forest landscape. We used null modeling to estimate the importance of selective and neutral assembly processes on community composition, and identified environmental factors influencing ecological selection through regression and structural equation modeling (SEM). Proportionally, the strongest process influencing community composition was dispersal limitation (0.36), exceeding the influence of homogenous selection (0.21), variable selection (0.16) and homogenizing dispersal (0.05). Fe(II) content was the most important factor explaining variable selection, and was significantly associated with total selection by univariate regression ($R^2 = 0.14$, $P = 0.003$). SEM supported a model in which Fe(II) content mediated influences of the Gibbs free energy of the organic matter pool and organic acid concentration on total selection. These findings suggest that the dominant processes shaping microbial communities in permafrost result from the stability of the permafrost environment, which imposes dispersal and thermodynamic constraints.

Received: 30 November 2017; **Accepted:** 13 May 2018

Keywords: permafrost; community assembly; null models; landscape microbial ecology; soil organic matter

INTRODUCTION

Permafrost is defined as ground that has remained below 0°C for two or more consecutive years (Van Everdingen 1998). Because this definition is solely based on a condition of 'ground climate' (Burn and Nelson 2006), permafrost-affected soils can span a diverse range of soil types and be highly varied in geography, geology, physicochemistry and microbiology. Indeed, permafrost environments account for approximately 16% of Earth's soil environments (Tarnocai *et al.* 2009), spanning much of the terrestrial Arctic and subarctic (Zhang *et al.* 2008), ice-free areas of Antarctica (Bockheim 1995) and high-elevation regions in both the northern and southern hemispheres (Zhang *et al.* 2008; Ping *et al.* 2015). Collectively, these soils represent an important microbial ecosystem (Jansson and Tas 2014) and a globally significant pool of sequestered carbon (Schuur *et al.* 2008; Tarnocai *et al.* 2009), which is being mobilized as climate warming increases permafrost thaw (Vaughan *et al.* 2013). While the fate of this carbon remains uncertain, it will likely be strongly dependent on properties of the resident microbial communities and the local soil conditions. As such, it is important to understand the natural abiotic and biotic variation that occurs within permafrost environments in order to accurately inform models aimed at predicting responses to change across these regions.

Descriptions of permafrost microbial communities have been reported from diverse environments and show a high level of variation in community composition. Studies of Arctic permafrost have described samples dominated by Chloroflexi, Proteobacteria and Actinobacteria from Hess Creek, Alaska (Mackelprang *et al.* 2011; Hultman *et al.* 2015), Bacteroidetes and Firmicutes from Hess Creek, Alaska (Mackelprang *et al.* 2011), Chloroflexi and AD3 in permafrost from Nome, Alaska (Tas *et al.* 2014), Proteobacteria and Actinobacteria in permafrost samples from the Canadian Arctic (Steven *et al.* 2007; Steven *et al.* 2008; Yergeau *et al.* 2010) and Bacteroidetes in permafrost alluvium from Eastern Siberia (Brouchkov *et al.* 2017). Community composition may be related to permafrost age and history, as recent analyses of Alaskan Pleistocene permafrost showed variation in community composition and metabolic capacity with respect to permafrost age over a chronosequence spanning 19–33kyr (Mackelprang *et al.* 2017). Analyses of Arctic submarine permafrost also show the bacterial community composition may be primarily influenced by sedimentation and thaw history, though permafrost conditions are dependent on the length of marine exposure (Mitzscherling *et al.* 2017). Permafrost origin is additionally important in driving community structure and function as has been reported from Siberian Pleistocene permafrost from lake-alluvial and Ice Complex (Yedoma) sediments, which showed differences in community composition and genetic potential for carbon, nitrogen and sulfur cycling between the two sites (Rivkina *et al.* 2016). Reports from the McMurdo Dry Valleys of Antarctica suggest that soil conditions may be additionally important for shaping the permafrost microbiome, as biomass and community composition of permafrost appear to vary between those found in Taylor Valley (Bakermans *et al.* 2014) and those found in the colder, drier, more oligotrophic soils of University Valley (Goordial *et al.* 2016).

While both environmental conditions and microbial community composition of permafrost-affected soils are known to be highly variable (Jansson and Tas 2014), the degree to which variation in community composition is linked to physicochemical conditions of the soil is not well understood (Mackelprang *et al.* 2016). In many non-permafrost soils, microbial community composition is shaped by physicochemical conditions, including pH (Fierer and Jackson 2006; Lauber *et al.* 2009; Chu *et al.* 2010), nutrient content (Cleveland *et al.* 2007; Ramirez, Craine and Fierer 2012) and soil moisture (Brockett, Prescott and Grayston 2012; Cruz-Martínez *et al.* 2012; Zhao *et al.* 2016). Given that permafrost can support active microbial communities (Steven *et al.* 2007; Tuorto *et al.* 2014), it is reasonable to assume that similar factors may be important in structuring the permafrost microbiome. Alternatively, microbial community composition in these environments may be decoupled from physicochemical conditions that are found to be important in non-permafrost soils, and may instead be shaped by the shared constraints imposed by prolonged freezing. Understanding how microbial communities are shaped by ecological processes and environmental conditions represents an important knowledge gap in permafrost microbiology.

While resolving drivers of community composition in permafrost environments will improve fundamental understanding of the microbiology of these extreme ecosystems, there is also practical importance in resolving how pre-thaw conditions may be used as predictors of system level response to thaw. Earth system models that integrate aspects of microbial community composition and function are gaining support to improve understanding of terrestrial carbon cycling and predictions about the fate of soil carbon in response to environmental change (Trivedi, Anderson and Singh 2013; Wieder *et al.* 2015). However, permafrost environments bring a high level of complexity that is difficult to generalize in current models, because soil type, soil conditions and carbon composition may all have important impacts on post-thaw dynamics and carbon transformation (Waldrop *et al.* 2010; Coolen *et al.* 2011; Lee *et al.* 2012; Ernakovich, Wallenstein and Calderón 2015). Additionally, the composition of pre-thaw communities may be a strong determinant of post-thaw processes, as permafrost microbial communities are expected to respond rapidly to thaw (Mackelprang *et al.* 2011; Coolen and Orsi 2015), and the abundance of particular taxa and functional genes can be important predictors of process rates, such as methanogenesis (Waldrop *et al.* 2010; McCalley *et al.* 2014; Mondav *et al.* 2014) and iron reduction (Hultman *et al.* 2015). These findings underscore the importance of integrating knowledge of the physical environment, the chemical nature of the organic matter pool and the structure and function of permafrost microbial communities to accurately predict rates of carbon metabolism in these systems. Spatially explicit studies capturing measures of soil heterogeneity are, therefore, necessary to inform models aimed at predicting microbial community responses to permafrost thaw and carbon fate in these environments.

The purpose of this work was to resolve the factors and processes that govern permafrost microbial community structure. We hypothesize that permafrost microbial communities are primarily shaped by the physical constraints imposed by prolonged freezing, and will therefore exhibit spatial distributions that reflect dispersal limitation and selective pressures associated with the physical constraints of the permafrost environment. To evaluate this hypothesis, we characterized patterns of permafrost microbial community composition along landscape gradients in the boreal forest ecosystem of the Caribou Poker Creek

Research Watershed (CPCRW) near Fairbanks, AK. We estimated the influence of variable selection (differences in taxa fitness with respect to conditions that vary spatially and that deterministically lead to heterogeneity in the ecology of taxa assembled into local communities), homogenous selection (differences in taxa fitness with respect to conditions that are homogenous spatially and that deterministically lead to homogeneity in the ecology of taxa assembled into local communities) homogenizing dispersal (high rates of organismal movement that lead to spatial homogeneity in the taxonomic profile of local communities) and dispersal limitation (low rates of organismal movement that lead to spatial heterogeneity in the taxonomic profile of local communities) (Dini-Andreote *et al.* 2015; Stegen *et al.* 2015). We further evaluated the importance of permafrost physicochemical conditions to impose deterministic selection on microbial communities. As the first landscape-scale survey linking permafrost community composition to environmental variability, this work provides mechanistic understanding of the controls on permafrost communities. This understanding can, in turn, inform models aimed at predicting permafrost microbial community characteristics and responses to thaw.

MATERIALS AND METHODS

Sample collection and processing

Permafrost samples were collected along a hydrologic gradient in the CPCRW in August, 2015. CPCRW is a long-term ecological research site and is representative of the discontinuous permafrost regions of interior Alaska (http://www.lter.uaf.edu/research/study-sites-cpcrw). The sampling site was located on a gentle southeast-facing slope (65.161°, −147.484°). To efficiently capture spatial variation at the landscape scale, we used a cyclic sampling design, as opposed to regular grid spacing (Burrows *et al.* 2002). A 3/5 cyclic sampling design with 4 m grid cells was employed along four replicate transects for 104 m: starting at the lowest elevation, transects were sampled at 0, 4, 12, 20, 24, 32, 40, 44, 52, 60, 64, 72, 80, 84, 92, 100 and 104 m. Four replicate transects ran parallel to each other based on a 2/3 cyclic sampling design with 10 m grid cells for 40 m (Fig. 1).

At each sampling position, permafrost depth was determined using a frost probe and permafrost cores were collected using a SIPRE coring auger (John's Machine Shop, Fairbanks, AK). Non-frozen soil material overlying ice-bonded permafrost was removed from the core barrel and discarded at the time of sampling, such that only ice-bonded permafrost samples were retained for analyses. Samples were wrapped in aluminum foil and packed in coolers on dry ice until they could be stored at −20°C at the University of Alaska, Fairbanks, AK. Samples were shipped on dry ice to Pacific Northwest National Laboratory in Richland, WA, where they were stored at −20°C until further processing.

The top 6 cm (SD = 2 cm) of each core was removed using an ethanol sterilized chisel to ensure removal of any material that may experience seasonal thaw and the next 7 cm (SD = 2 cm) section of each core was taken for analysis. These values vary slightly between cores due to practical constraints and properties of individual core samples (length of the original core samples obtained and natural fracture points that existed in several cores). The length of material removed from the top of each sample and the length of the core kept for analysis is presented in Table S1 (Supporting Information). The exterior of each core section was removed using sterile razor blades, starting from a pristine surface of the core. Decontaminated cores were crushed and homogenized while frozen using a sterile stainless steel soil press in a walk-in −20°C freezer. Homogenized frozen samples were partitioned aseptically for downstream analyses. Samples to be stored anaerobically were immediately purged with nitrogen gas in 60 ml serum bottles sealed with butyl rubber stoppers, and all partitioned material was stored at −20°C until analysis. A total of 59 permafrost samples were included in the final analyses, as cores could not be retrieved from some sample locations or were compromised during sample processing (refer to Fig. 1).

Physicochemical analyses

Soil water content was determined by drying 1–10 g of sample at 105°C, and measuring mass loss after 48 h: sample masses were determined after 24 and 48 h to ensure samples reached a constant mass in consecutive measurements. Water content was determined from the average of five replicate measurements per sample.

Total carbon and nitrogen content was determined from 30 mg freeze-dried, ground and <2 mm sieved samples. Samples were analyzed on an Elementar vario El cube (Elementar, Germany). Values were determined from the average of triplicate measurements for each sample.

Samples for metals and anion analyses were prepared from freeze-dried, ground and <2 mm sieved samples. For metals analysis, 1 g of sample was extracted with 10 ml of 0.5 N HCl shaking at 200 rpm for 2 h at room temperature. Anion extractions were completed as above, with 1 g of soil in 5–10 ml of deionized water. Metals (Fe, Mn, Mg, Cu, P, S) and anion (Cl^-, SO_4^-, and NO_3^-) analyses were completed as previously described (Zachara *et al.* 2016).

Iron(II) content was determined by ferrozine assay (Stookey 1970). In an anaerobic chamber (Coy Laboratory Products, Grass Valley, MI), 10 ml 0.5 N HCl was added to 1 g of permafrost sample, and the vial was sealed and vortexed. Samples were extracted for 1 h and filtered through a 0.22 μm pore-size polyethersulfone syringe filter. Extracts were diluted in 0.1 N HCl and 100 μl was added to 1 ml ferrozine; after 5 min, the absorbance at 562 nm was measured on a Shimadzu Biospec-1601 spectrophotometer. Iron(II) concentrations were determined from a six-point standard curve ranging from 0 to 45 μM Iron(II). Samples were dried at 60°C and weighed to determine the Iron(II) content by dry weight.

Organic acids and sugars were quantified in the same water extracts prepared for anion analysis, using an Agilent 1100 series HPLC (Agilent, Palo Alto, CA) with a 300 × 7.8 mm Aminex HPX-87H column (Bio-Rad, Hercules, CA), a 0.008 N H_2SO_4 mobile phase with a flow rate of 0.6 ml/min and variable wavelength detector at 210 nm for organic acids and refraction index detector for sugars. Samples were filtered through a 0.22 μm pore-size polyethersulfone syringe filter and acidified by adding 10 μl of 2.5 N H_2SO_4 per ml. Concentrations were determined by comparison to peak areas of standards.

Soil texture was determined by measuring the gravel (>2 mm), sand (64 μm–2mm), and mud (silt and clay) (<64 μm) fractions of each sample. Briefly, 20 g of soil was dried at 60°C, and the total dry weight determined. Samples were dry sieved through a 2 mm sieve, and the mass of the >2mm fraction was used to calculate the gravel fraction. The <2 mm fraction was wet sieved through a 64 μm sieve, the fraction retained was dried and used to calculate the sand fraction, while the <64 μm fraction was dried and used to calculate the mud fraction.

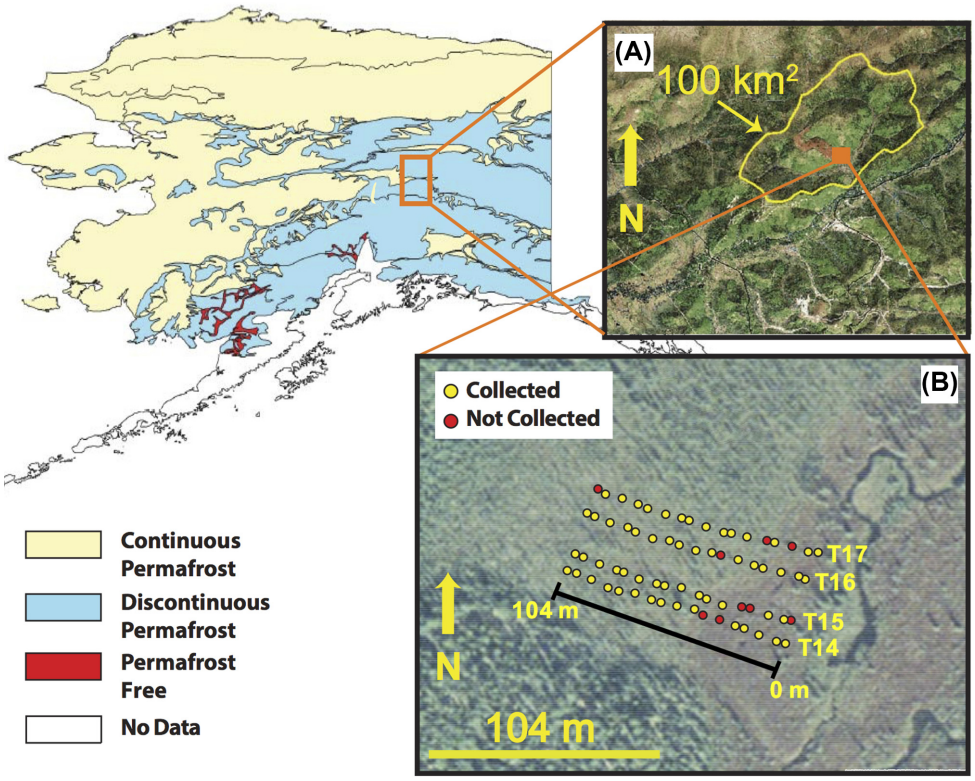

Figure 1. Map of Alaska indicating the location of **(A)** the Caribou Poker Creek Research Watershed (CPCRW) and **(B)** the location of the sample sites along each transect. Yellow dots indicate where samples were taken and included in the final analysis, while red dots indicate landscape positions where samples could not be recovered or where samples were compromised during processing, such that they were excluded from the final analysis.

Soil pH was determined on a Denver Instrument model 215 pH meter (Denver Instruments, Bohemia, NY) by slurry of 1 g soil in 2 ml MilliQ water (Millipore Sigma, St. Louis, MO).

Organic matter composition determination by FT-ICR-MS

Organic matter was extracted from bulk soil sequentially with water, methanol and chloroform as described previously (Tfaily *et al.* 2015, 2017). Briefly, organic matter extracts were prepared by adding 1 ml of solvent to 100 mg lyophilized and ground bulk soil and shaked for 2 h on an Eppendorf Thermomixer in 2 ml capped glass vials. Samples were removed from the shaker and left to stand before centrifugation at 2000 rpm for 10 min and the supernatant was retained for analysis. The soil residue was dried with nitrogen gas to remove any residual solvent, and the extraction was repeated with each of the next two solvents. The chloroform and water extracts were diluted in methanol to improve electrospray ionization (ESI) efficiency and 20 μl was injected into the FTICR-MS. Samples were analyzed in triplicate for water extractions and chloroform extractions, and singly for methanol extractions.

A 12 Tesla Bruker SolariX FTICR-MS located at the Environmental Molecular Sciences Laboratory in Richland, WA, was used to collect high-resolution mass spectra of the organic matter in the extracts. A standard Bruker ESI source was used to generate negatively charged molecular ions. Samples were introduced directly to the ESI source at a flow rate of 3 μl/min. The ion accumulation time was varied, from 0.1 to 0.5 s, to account for differences in C concentration between samples and to maintain a final dissolved organic carbon concentration of 20 ppm. The instrument was externally calibrated weekly with a tuning solution from Agilent (Santa Clara, CA), which calibrates to a mass accuracy of <0.1 ppm. Two hundred scans were averaged for each sample and internally calibrated using OM homologous series separated by 14 Da ($-CH_2$ groups). The mass measurement accuracy was less than 1 ppm for singly charged ions across a broad m/z range (i.e. 200 < m/z <1200). To further reduce cumulative errors, all sample peak lists for the entire dataset were aligned to each other prior to formula assignment to eliminate possible mass shifts that would impact formula assignment. Putative chemical formulas were assigned using in-house software based on the Compound Identification Algorithm (Kujawinski and Behn 2006), and modified as previously described (Minor *et al.* 2012). Chemical formulas were assigned based on the following criteria: S/N > 7, and mass measurement error <1 ppm, taking into consideration the presence of C, H, O, N, S and P and excluding other elements. Peaks with large mass ratios (m/z values > 500 Da) were assigned formulas through the detection of homologous series (CH_2, O, H_2). Additionally, to ensure consistent assignment of molecular formula the following rules were implemented: one phosphorus requires at least four oxygens in a formula and when multiple formula candidates were assigned the formula with the lowest error and with the lowest number of heteroatoms was picked.

For all analyses, peak intensities were converted to presence/absence and peaks observed in any of the triplicate measurements were included as present. Compound classes were assigned to chemical formulas based on molar O:C and H:C ratios, determined from analysis of van Krevelen diagrams. The Gibbs energies of the oxidation half-reaction (ΔG°_{Cox}) of each compound was derived based on the nominal oxidation state

of carbon as previously described (LaRowe and Van Cappellen 2011). The average ΔG°_{Cox} of the carbon pool was determined for each sample extraction: the median values were used for the methanol ($\Delta G^{\circ}_{Cox}(MeOH)$) and chloroform ($\Delta G^{\circ}_{Cox}(CHCl_3)$) extracts due to highly skewed distributions, while the ΔG°_{Cox} was normally distributed for water extracts ($\Delta G^{\circ}_{Cox}(H_2O)$) such that the mean values were used.

Microbial community analyses

Total community DNA was extracted from 0.25 g of each sample using the MoBio Power Soil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA), according to manufacturer's instructions. Additional cleanup and concentration of DNA extracts was completed using the Zymo ZR-96 Genomic DNA Clean and Concentrator-5 kit (Zymo Research Corporation, Irvine, CA). PCR amplification of the V4 region of the 16S rRNA gene was performed as previously described (Caporaso *et al.* 2012), with the exception that the 12-base barcode sequence was included in the forward primer. Amplicons were sequenced on an Illumina MiSeq using the 500 cycle Miseq Reagent Kit v2 (Illumina Inc., San Diego, CA), according to manufacturer's instructions.

Sequence processing was completed using default parameters of the following tools, except where stated. Raw sequence reads were demultiplexed using EA-Utils (Aronesty 2013) not allowing any mismatches in the barcode sequence. Reads were quality filtered with BBDuk2 (Bushnell 2014) to remove adapter sequences and PhiX with matching kmer length of 31 bp at a hamming distance of 1. Reads shorter than 51 bp were discarded. Reads were merged using USEARCH (Edgar 2010) with a minimum length threshold of 175 bp and maximum error rate of 1%. Sequences were de-replicated and clustered using the distance-based, greedy clustering method of USEARCH at 97% pairwise sequence identity among operational taxonomic unit (OTU) member sequences. Taxonomy was assigned to OTU sequences at a minimum identity cutoff fraction of 0.8 using the global alignment method implemented in USEARCH across RDP trainset version 15. OTU seed sequences were filtered against RDP classifier training database version 9 to identify chimeric OTUs using USEARCH. De novo prediction of chimeric reads occurred as reads were assigned to OTUs. OTU count tables were randomly resampled to 17 899 sequences and OTUs that could not be assigned at the kingdom level were removed.

Sequence data has been deposited in the European Nucleotide Archive (ENA), under accession number PRJEB23054 (http://www.ebi.ac.uk/ena/data/view/PRJEB23054).

Statistical analysis

The environmental variables consisted of all physicochemical variables and the average ΔG°_{Cox} for each FTICR extraction (mean for $\Delta G^{\circ}_{Cox}(H_2O)$ and median for $\Delta G^{\circ}_{Cox}(MeOH)$ and $\Delta G^{\circ}_{Cox}(CHCl_3)$). Missing data were replaced by the geometric mean of values for a given variable, or the arithmetic mean in the case of the lactate data, which had numerous zero values. Data for water content, Cl, SO_4, NO_3, Fe(total), Mn, Mg, Cu, P, S, Fe(II), C and N were $\log_{10}(x)$ transformed, and data for lactate, formate and acetate concentrations were $\log_{10}(x+1)$ transformed. Data for pH, gravel, sand, mud, $\Delta G^{\circ}_{Cox}(H_2O)$, $\Delta G^{\circ}_{Cox}(CHCl_3)$ and $\Delta G^{\circ}_{Cox}(MeOH)$ were not transformed.

Principal components analysis (PCA) was used to assess variation in environmental variables across the landscape using the *princomp* function in R (RStudioTeam 2015). All variables were scaled by subtracting the mean and dividing by the standard deviation prior to analysis. Scores of all principal components (PCs) and variable loadings along each PC were extracted for downstream analyses. Variable loadings along each PC were used to assess the importance of individual variables to each PC.

Analyses of community diversity and composition were completed using the '*vegan*' package (Oksanen *et al.* 2017) in R. Shannon diversity estimates were completed based on the resampled OTU counts. OTU abundances were Hellinger transformed prior to all other compositional analyses. Non-metric multidimensional scaling was used to examine the community variation between samples based on Bray Curtis dissimilarity, and environmental variables were fit as vectors in the final two-dimensional ordination to evaluate relationships between community and environmental variation.

Spatial analyses were completed in R. Kriging was used to interpolate and visualize spatial trends in both the environmental and biological data, using the autokrig function of the '*automap*' package (Hiemstra *et al.* 2009). Principal coordinates of neighbor matrices (PCNM) were used to create orthogonal spatial variables based on sample site locations (Borcard and Legendre 2002; Borcard *et al.* 2004). PCNMs were calculated as previously described (Borcard, Gillet and Legendre 2011). PCNM axes were used as explanatory variables in downstream analyses to examine the importance of spatial filters on community composition.

A redundancy analysis (RDA) model was used to relate community composition to environmental and spatial variation using the *vegan* package (Oksanen *et al.* 2017) in R. Due to collinearity between several environmental variables, PC scores extracted from the environmental PCA were used to represent environmental variables in the model. Forward stepwise model building based on adjusted R^2 was carried out using all 23 PCs, all 15 positive PCNMs and all variables related to sample depth and sample size listed in Table S1 (Supporting Information) (active layer depth, total depth to the top of the processed sample, total depth to the bottom of the processed sample and total length of the core sample analyzed). The importance of each variable added to the model was assessed using variance partitioning based on RDA.

Null modeling was used to estimate the influence of ecological processes on community composition, as described previously (Dini-Andreote *et al.* 2015; Stegen *et al.* 2015). The influence of selection was estimated by evaluating the difference between the observed between-community mean-nearest-taxon distance (βMNTD) (taxon was defined at the OTU level) and the mean of the null distribution in units of standard deviation. Significant deviations from the null distribution were evaluated using the β-nearest taxon index (βNTI) and the signal for selection was expressed as the proportion of comparisons for which $\beta NTI > 2$ or $\beta NTI < -2$, representing signals for variable selection and homogenous selection, respectively. Comparisons falling within the null distribution ($2 > \beta NTI > -2$) represent compositional differences that do not arise from selection, and are instead attributable to dispersal limitation, homogenizing dispersal, or processes undominated by dispersal or selection. To assess the relative influence of these processes, a Raup-Crick metric incorporating species relative abundance (RC_{bray}) was used to compare the observed and expected species turnover between communities. Significant deviations from the null distribution indicating greater than expected differences in community composition ($2 > \beta NTI > -2$ and $RC_{bray} > 0.95$) were attributed to dispersal limitation, while those indicating less than expected differences in community composition ($2 > \beta NTI$

> −2 and RC_{bray} < −0.95) were attributed to homogenizing dispersal. Comparisons falling within the null distribution of both metrics (2 > βNTI > −2 and 0.95 > RC_{bray} > −0.95) represent differences in community composition that were not strongly governed by selection or dispersal (i.e., the observed differences were 'undominated').

A regression modelling approach was used to identify the environmental variables that explain variation in the process estimates for total selection (variable and homogenous selection combined). Here, process estimates were generated for each community by finding the fraction of pairwise comparisons—between a given community and all other communities—falling into the process categories summarized above (Stegen *et al.* 2015). Community-level estimates of total selection were then used as the dependent variable in an exhaustive model selection using Bayesian information criterion, performed in the *'leaps'* package (Lumley 2017) in R.

Path analysis was used to estimate interactions among environmental variables predicted to influence total selection. A hypothetical model outlining expected relationships between variables was evaluated using the sem function of the *'sem'* package (Fox, Zhenghua and Byrnes 2017) in R (Fig. S1, Supporting Information). Adjustments to the model were informed by modification indices, which suggest addition of paths to improve model fit, and were included based on logical evaluation of potential associations between variables.

RESULTS

Environmental conditions and carbon composition

Permafrost characteristics were highly variable across the sampling area, and are summarized in Table S2 (Supporting Information). Samples ranged greatly in carbon content from 1.3% to 35.8%, and nitrogen content co-varied strongly with carbon content (R^2 = 0.98), ranging from 0.1% to 2.0%. Soil texture was typically dominated by sand (average 63%), but had substantial inputs of mud (average 35%). All samples were mildly acidic, ranging from pH 4.9 to 6.7. Notably, permafrost samples across the site varied greatly in ice content, with gravimetric water content varying from 0.28 to 9.2 g(water)/g(dry soil). Fe(II) content, indicative of soil redox conditions, was also highly variable, spanning over two orders of magnitude from 0.07 to 12.9 mg/g(dry soil).

The compound classes assigned to FTICR peaks based on van Krevelen diagrams showed distinct peak profiles for each solvent extraction (Tables S3–S6, Supporting Information). Water extractions recovered the highest percentage of compounds classified as lignin-, condensed hydrocarbon-, carbohydrate-, tannin- and amino sugar-like compounds, while methanol and chloroform extractions recovered the highest percentage of compounds grouping to unsaturated hydrocarbon- and lipid-like compounds. Compounds grouping as peptide- or protein-like were recovered in all fractions, representing 6.6% (SD = 1.4%), 8.9% (SD = 2.9%) and 5.0% (SD = 0.9%) in the water, methanol and chloroform extracts, respectively. A large percentage of compounds in each extraction were not assigned to a compound class (25%–48%). The ΔG°_{Cox} estimates from the FTICR profiles were tightly linked to the overall variation in FTICR compound classes for each extraction (Fig. S2, Supporting Information). The ΔG°_{Cox} estimates were, therefore, used to represent organic carbon profiles in downstream analyses, as they capture variation in organic carbon composition as a biochemically meaningful continuous variable that can be interpreted mechanistically.

PCA using all physicochemical variables revealed environmental gradients both within and between transects (Fig. 2). The first two PCs accounted for nearly 58% of the environmental variance, with 41% captured on PC1 and 17% on PC2. The strongest loadings along PC1 were for C content (−0.31), N content (−0.31), water content (−0.28), S content (−0.28), acetate (−0.28), $\Delta G^{\circ}_{Cox}(H_2O)$ (−0.27) and formate (−0.27), while the strongest loadings along PC2 were for Fe(II) content (−0.37), soil texture fractions of mud (−0.37) and sand (0.35), pH (0.33) and P content (−0.33).

Microbial community composition

Bacterial sequences grouped to a total of 45 phyla or candidate phyla, and 11 phyla were represented at >1% of the total community (Fig. S3, Supporting Information). Based on the average number of sequences in each sample grouping to bacterial phyla, communities were dominated by Proteobacteria (24%, SD = 3%) (particularly Beta- (11%, SD = 2%), Alpha-(6%, SD = 2%), Delta- (5%, SD = 2%), and Gamma- (1%, SD = 1%) proteobacteria), Acidobacteria (17%, SD = 5%) (particularly subgroup 7 (4%, SD = 2%), subgroup 4 (4%, SD = 4%), subgroup 6 (2%, SD = 1%), subgroup 3 (2%, SD = 1%), and subgroup 1 (2%, SD = 1%)), Verrucomicrobia (13%, SD = 4%), Actinobacteria (10%, SD = 3%), Chloroflexi (10%, SD = 3%), Bacteroidetes (8%, SD = 5%), Gemmatimonadetes (5%, SD = 2%), Planctomycetes (2%, SD = 1%), Nitrospirae (2%, SD = 1%), Parcubacteria (1%, SD = 1%) and Firmicutes (1%, SD = 1%). Bacterial sequences grouping to other phyla and bacterial sequences that could not be classified at the phylum level represented 5% (SD = 2%) and 1% (SD = 1%) of the total sequences, respectively.

Approximately 1% of sequences were classified as Archaeal, with 79% of these sequences grouping to the phylum Euryarchaeota. Sequences within the Euryarchaeota grouped predominantly within the Methanomicrobia and Methanobacteria.

Patterns of community composition

Community composition showed non-random spatial structure (Fig. 3), and was explained by both environmental variables (PCs) and, to a lesser extent, spatial variables (PCNMs). Stepwise model selection supported a model with 15 variables, which fit the data with an adjusted R^2 = 0.48; however, variance partitioning showed that many of these variables contributed only incrementally to improving model fit (Fig. S4, Supporting Information). A model incorporating the first two variables from the selected model (PC2, PC1) fit the data with an adjusted R^2 = 0.34, and subsequent addition of the remaining variables retained in the selected model improved the adjusted R^2 by 0.02 (PCNM4) or less (all other variables) (Fig. 4). We therefore focused our interpretation on the model including PC2, PC1 and PCNM4. The variable loadings on the PCs selected in the model indicated that several environmental factors were related to community composition (see Fig. 2 for relationships between environmental factors along PC1 and PC2).

Univariate regression of factors with the strongest loadings along PC1 and PC2 showed that alpha diversity and the relative abundances of particular taxa were significantly associated with one or more of these environmental variables (Table S7, Supporting Information).

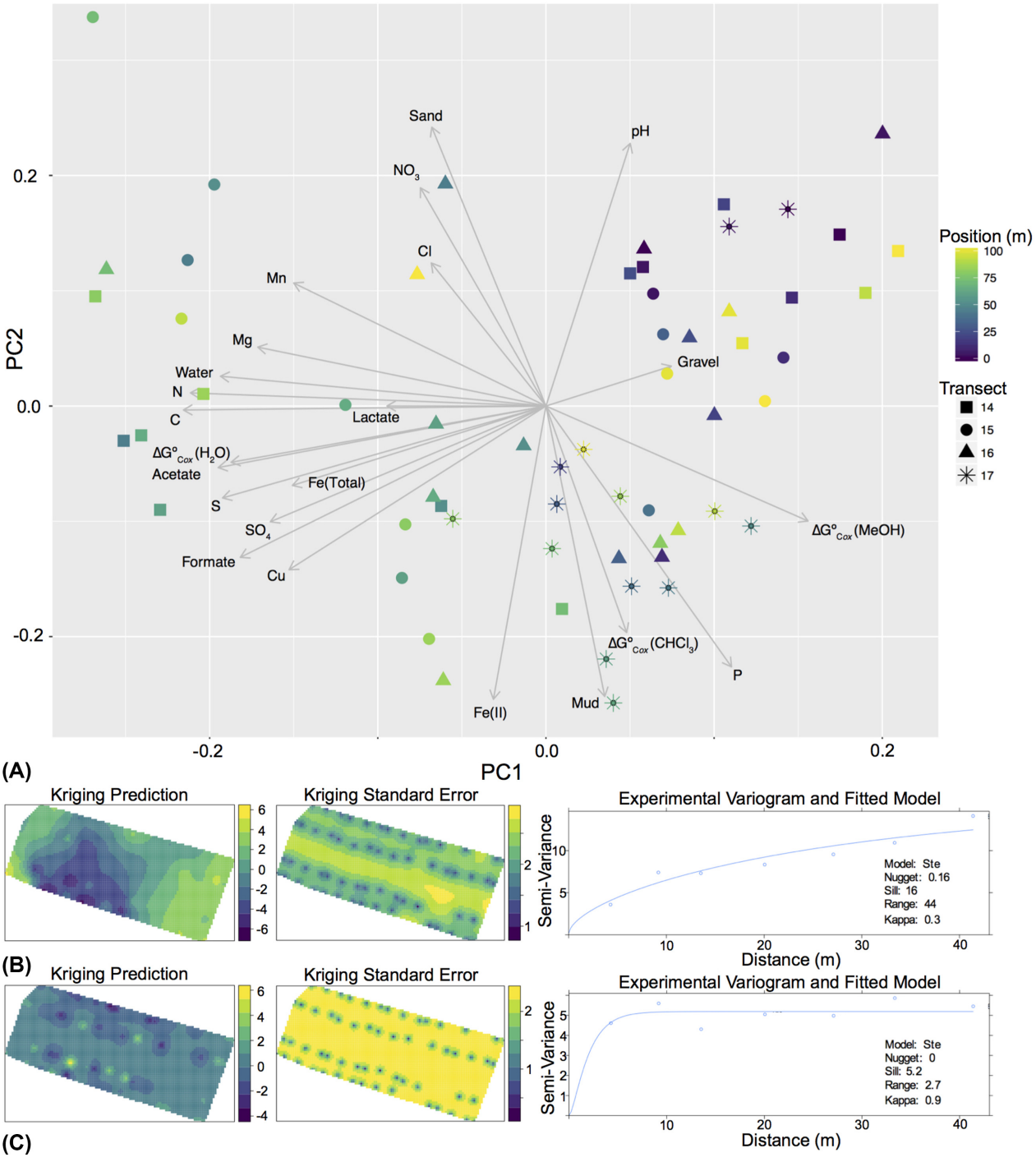

Figure 2. **(A)** Principal components analysis (PCA) representing environmental variation between samples, and Kriging predictions of spatial patterns of environmental variation based on **(B)** PC1 and **(C)** PC2 scores across the sampling area (n = 59).

Null model analyses

Null modeling revealed signals for variable selection (βNTI > 2), homogenous selection (βNTI < −2), dispersal limitation (2 > βNTI > −2 and RC_{bray} > 0.95), homogenizing dispersal (2 > βNTI > −2 and RC_{bray} < −0.95) and processes undominated by dispersal or selection (2 > βNTI > −2 and 0.95 > RC_{bray} > −0.95) (Fig. 5). Values from 0 to 1 indicating the relative influence of each process on the observed variation in community composition revealed that the strongest signal was for dispersal limitation (0.36), and the lowest signal was for homogenizing dispersal (0.05). The signal for homogenous selection (0.21) was slightly higher than for variable selection (0.16), contributing to a signal of 0.37 for total selection. Variation not accounted for by dispersal or selection accounted for the remaining signal of 0.23.

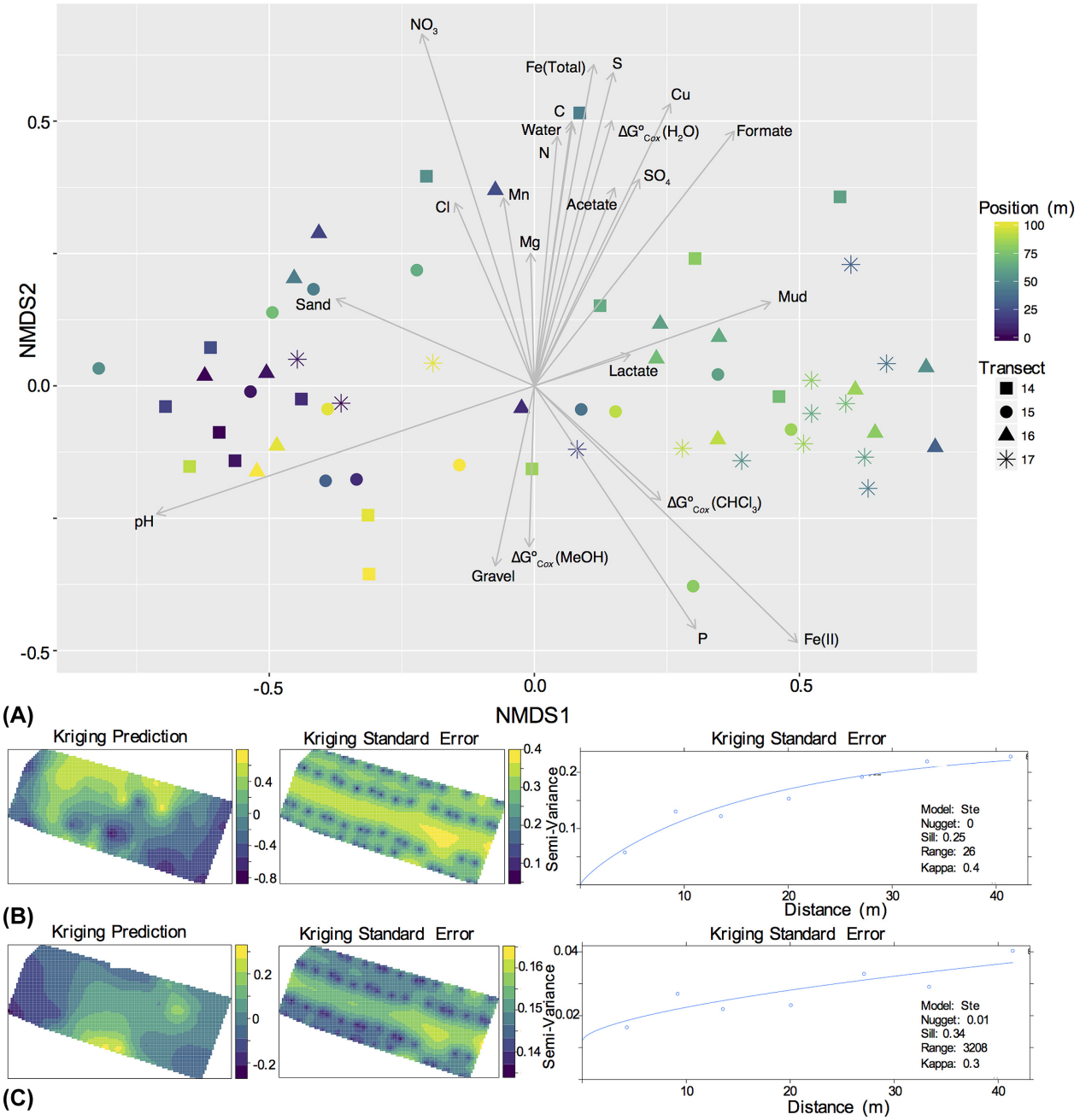

Figure 3. **(A)** Non-metric multidimensional scaling (NMDS) plot representing the Bray Curtis dissimilarity in microbial community composition between samples, with environmental vectors overlaid, and Kriging predictions of spatial patterns of community composition based on **(B)** NMDS1 and **(C)** NMDS2 scores across the sampling area (n = 59).

Regression model selection indicated Fe(II) was the most important environmental variable influencing variable selection, and Fe(II) was significantly associated with the relative influence of total selection by univariate regression ($R^2 = 0.14$, $P = 0.003$).

Path analysis

Given the relationship observed between Fe(II) and total selection, we proposed a path model in which variables that reflect energetic constraints on microbial activity may influence total selection indirectly, through relationships mediated by Fe(II) content (see Fig. S1, Supporting Information). Our initial model was not consistent with the data ($X^2 = 37.2$, d.f. = 9, $P = 2.4\times10^{-5}$), and was revised to better reflect relationships between the variables. All paths in the initial model were retained in the final model, and modification indices supported the addition of a path from soil carbon content to nitrate content. The final model did not differ significantly from the data ($X^2 = 10.3$, d.f. = 8, $P = 0.25$) and explained 14.5% of the variation in total selection and between 37% and 68% of the variation in other endogenous variables (Fig. 6). The direct effect of Fe(II) was the strongest total effect on total selection,

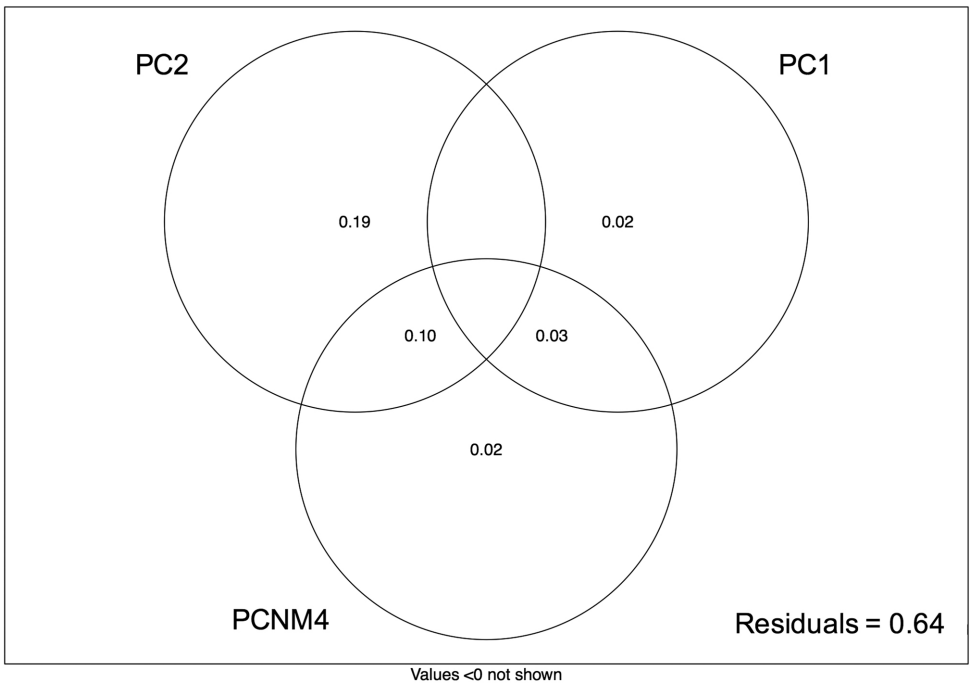

Figure 4. The proportion of variation in microbial community composition explained by the environmental and spatial variables selected in forward stepwise model building (n = 59): including additional variables improved the adjusted R^2 of the model by <0.02.

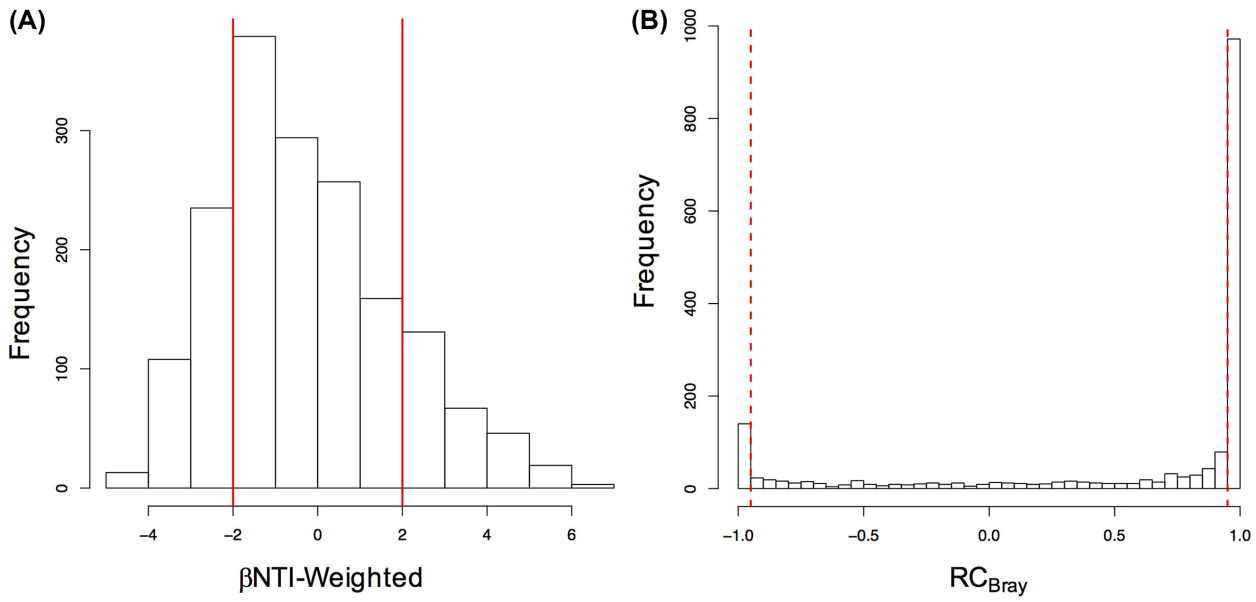

Figure 5. Histograms representing the observed distribution of comparisons based on βNTI and RC_{Bray}. Red lines represent the significance thresholds, whereby values outside their bounds are significantly different from the null distribution (n = 59).

while organic acid content had the strongest indirect effect on total selection (Table 1).

DISCUSSION

Previous studies have demonstrated that permafrost soils contain diverse and varied communities (reviewed in Jansson and Tas 2014) that are likely active *in situ* (Steven *et al.* 2007; Tuorto *et al.* 2014; Mackelprang *et al.* 2017); however, studies designed to examine ecological processes that may influence patterns of community structure and function in permafrost are lacking. To address this knowledge gap, we characterized patterns of environmental variation and permafrost microbial community composition in a boreal forest ecosystem across landscape gradients. By employing a well-replicated and geostatistically-informed sampling design, we have provided the first characterization of ecological processes driving landscape scale spatial structure of permafrost microbial community composition. Through this work, we show that patterns of both environmental characteristics and microbial community composition can be highly variable over short distances and exhibit non-random spatial structure, with patterns of community composition

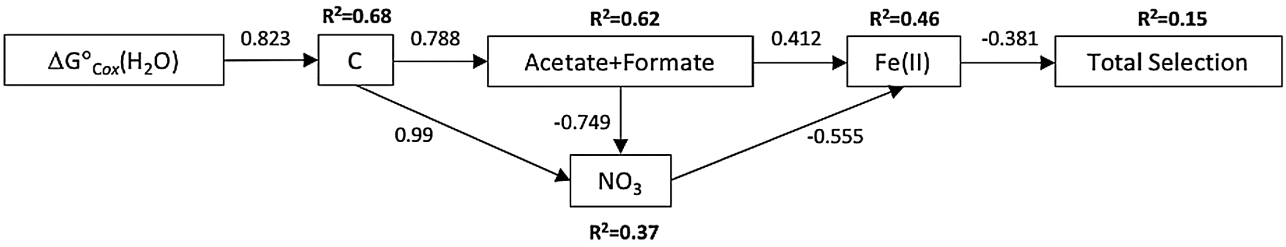

Figure 6. Final structural equation model ($X^2 = 10.3$, d.f. = 8, $P = 0.25$) representing relationships between variables hypothesized to deterministically influence community composition (n = 59). Values alongside arrows represent standardized path coefficients, and the variation explained for endogenous variables is indicated above each variable. All paths are significant.

Table 1. Standardized direct effects, indirect effects and total effects of environmental factors on total selection.

	Direct effects	Indirect effects	Total effects
$\Delta G^{\circ}_{Cox}(H_2O)$		−0.023	−0.023
Carbon		−0.039	−0.039
Acetate+Formate		−0.315	−0.315
Nitrate		0.211	0.211
Fe(II)	−0.381		−0.381

driven by deterministic and neutral processes that likely arise primarily from the physical constraints of the permafrost environment. We additionally suggest that the patterns we observed arise from contemporary active processes influencing community composition in the permafrost, specifically as a strong signal of dispersal limitation suggests that microbes have been reproducing and dying at rates that allow community composition to drift apart due to spatial isolation between locations.

Spatial structure of permafrost physicochemistry

The degree of heterogeneity in soil physicochemical and organic matter characteristics observed over the study area was striking, likely reflecting spatially structured variation in thaw history and organic matter deposition. We observed a non-linear spatial trend in environmental variation, with samples at the extreme ends of the transects found to be more similar to each other than to those in the middle of the transects, most notably in terms of water content, pH, carbon and nitrogen content, $\Delta G^{\circ}_{Cox}(H_2O)$ and organic acid content (acetate and formate). Higher water content, which was observed predominantly through the middle of the transects, may represent ice inclusions in the transition zone near the surface of the permafrost, formed during more recent thaw events (Shur, Hinkel and Nelson 2005). Total carbon and nitrogen content, $\Delta G^{\circ}_{Cox}(H_2O)$ and the abundance of organic acids were also highest through the middle of the transects, which may reflect more substantial deposition of undecomposed plant material from the active layer into the upper permafrost. The proportion of organic compounds grouping to lignins, carbohydrates and amino sugars was highest in water extracts from the middle of the transects, and substantial deposits of fibric material were observed in many of these same samples. This undecomposed plant matter likely contributes high ΔG°_{Cox} compounds, such as lignin-like compounds, increasing the average ΔG°_{Cox} of the carbon pool. This non-linear trend in environmental variation was not expected based on characteristics of the local topography.

We suggest that the higher organic acid concentrations observed in the middle of the spatial domain arise from the fermentation of labile organic compounds derived from deposited plant matter. If the most thermodynamically favorable compounds are preferentially fermented, this would further increase the average ΔG°_{Cox}. In sediments, a net accumulation of organic acids is observed when fermentation rates exceed respiration rates (McMahon and Chapelle 1991), and acetate and C_1 compounds are the dominant organic products of anaerobic metabolism in northern wetlands and bogs (Hines, Duddleston and Kiene 2001; Duddleston *et al.* 2002). These products of anaerobic metabolism may accumulate in permafrost through equivalent processes.

Microbial community composition and environmental correlates

Community composition across our study site shared similarities with permafrost communities reported previously from across the Arctic. We observed high representation of Proteobacteria (24%, SD = 3%), Acidobacteria (17%, SD = 5%), and Chloroflexi (10%, SD = 3%), which is largely consistent with reports of other Alaskan Holocene permafrost samples (Mackelprang *et al.* 2011; Tas *et al.* 2014; Hultman *et al.* 2015). Archaeal communities represented only a small percentage (1%) of the libraries and were dominated by taxa grouping to methanogens in the phylum Euryarchaeota, which is consistent with previous reports from across the Arctic (Mackelprang *et al.* 2011; Tas *et al.* 2014; Hultman *et al.* 2015; Rivkina *et al.* 2016). In contrast with previous studies, we saw high representation of Verrucomicrobia (13%, SD = 4%), which are globally abundant in soils (Bergmann *et al.* 2011), but have not been previously reported as dominant members of permafrost communities (Steven *et al.* 2007, 2008; Yergeau *et al.* 2010; Mackelprang *et al.* 2011; Tas *et al.* 2014; Hultman *et al.* 2015; Rivkina *et al.* 2016; Brouchkov *et al.* 2017). Further comparisons of geographically distinct permafrost communities will require an increased number of studies employing well-replicated sampling designs and the adoption of standardized analytical techniques within the field (Jansson and Tas 2014; Mackelprang *et al.* 2016).

Permafrost communities across the study site were influenced by similar drivers to those that have been reported previously for non-permafrost soil communities (Fierer and Jackson 2006; Lauber *et al.* 2009; Chu *et al.* 2010), however several relationships may be indicative of the unique constraints of the permafrost environment. Diversity was best described by a positive linear relationship with pH, which is consistent with trends observed in non-permafrost soils (Fierer and Jackson 2006; Lauber *et al.* 2009; Chu *et al.* 2010). The overall variation in

community composition showed a clear relationship with environmental variation (PCs), although the particular environmental variables influencing community variation were not clear from the RDA model selection. No strong relationship between pH and community composition was observed, perhaps due to the relatively small range in soil pH across samples (pH 4.9–6.7). The relative abundance of the dominant phyla also varied significantly with numerous environmental variables, however these relationships were atypical of trends observed in surveys of non-permafrost soils. For example, at the phylum level, Acidobacteria are typically negatively associated with pH, while Bacteroidetes and Actinobacteria typically have positive relationships with pH (Lauber *et al.* 2009); however, we observed the opposite trends for both Acidobacteria and Bacteroidetes and no trend for Actinobacteria with soil pH. Selective constraints of the permafrost environment may limit the phylogenetic breadth of these taxa, altering phylum level trends from those observed in other soils. For example, the Acidobacteria were dominated by subgroup 7 (4%, SD = 2%), subgroup 4 (4%, SD = 4%) and subgroup 6 (2%, SD = 1%), which typically show the positive relationships with soil pH that we observed (Jones *et al.* 2009). Additionally, other deterministic factors, such as soil redox conditions and soil organic matter composition, which also showed strong univariate relationships with the relative abundance of particular taxa, may be more important drivers of community structure in permafrost-affected soils.

Ecological processes influencing community composition

We employed a null modelling approach to evaluate the degree to which deterministic processes drive community variation and to resolve the variables most likely to be causally influencing composition. Patterns of community composition arise from a combination of deterministic and stochastic process (Vellend 2010) and the relative importance of these processes vary between systems. Null modeling provides a valuable tool to disentangle the influence of individual processes on patterns of microbial distribution (Stegen *et al.* 2013, 2015). This approach has significant advantages over the RDA models, which cannot estimate relative contributions of assembly processes and did not reveal specific environmental variables that drive spatial variation in community composition.

Null modeling revealed a strong signal for dispersal limitation combined with a very weak signal for homogenizing dispersal, indicative of very restricted movement of microorganisms within the permafrost. The signal for dispersal limitation was stronger than for either homogenous or variable selection, and was effectively equivalent to the value for total selection. Dispersal limitation may be an especially important process in permafrost-affected soils, where microorganisms remain frozen in place for prolonged periods. Significant dispersal events may therefore be restricted to the limited movement that occurs through cryoturbation. These constraints likely limit community mixing over very short distances, which would lead to the strong signal of dispersal limitation observed in our null model analyses.

Given strong dispersal limitation, we expected that the spatial PCNM variables would explain significant variation in community composition, independent of environmental variation. This expectation was not met, however, with PCNM axes explaining little variation in the RDA model. The lack of a strong spatial signal in the RDA model indicates that the influence of spatial processes manifest below the spatial resolution of our sampling, consistent with very restricted movement of microorganisms through permafrost. This is in contrast to other landscape scale studies that observe variation in community composition over much larger spatial scales (km) and attribute community variation predominantly to environmental heterogeneity over the landscape (Griffiths *et al.* 2011; Lee *et al.* 2012).

We found that 37% of the total community variation in permafrost community composition was explained by selective processes, and that soil characteristics associated with Fe(II) content are likely the most important environmental variables deterministically influencing community composition. Soil Fe(II) accumulates in anaerobic soils through the reduction of Fe(III), and iron reduction can contribute substantially to respiration in Arctic soils (Lipson *et al.* 2010, 2013). A recent multi-omic analysis of Alaskan permafrost reported high representation of proteins annotated to iron-reducing taxa and the expression of genes annotated as cytochromes central to iron-reduction, suggesting iron-reducing taxa were likely active *in situ* (Hultman *et al.* 2015). Importantly, Fe(III) reduction competes with other anaerobic processes, and suppresses less thermodynamically favorable methanogenic pathways (Miller *et al.* 2015). The relationship between Fe(II) and total selection indicates that soil redox conditions and thermodynamic constraints on microbial metabolism are likely to be the primary selection pressures that deterministically govern community composition.

These findings suggest that the stability of the permafrost environment strongly influences community structure and function, directly by restricting community mixing and indirectly by influencing the selective landscape, as electron donors and acceptors are depleted and infrequently replenished. This contrasts with non-permafrost soils, in which communities are presumed to be well-dispersed through aeolian (Barberán *et al.* 2015) and hydrologic process (Treves *et al.* 2003), nutrient fluxes are dynamic (Park and Matzner 2003; Austin *et al.* 2004), and communities are thought to be shaped predominantly by selection (Fierer, Bradford and Jackson 2007; Hartmann *et al.* 2009). Permafrost community structure and function, therefore, appear to be uniquely influenced by a balance between dispersal limitation imposed by frozen soil and deterministic selection arising primarily from thermodynamic constraints.

Thermodynamic constraints

While our final SEM model is most appropriately interpreted as a multi-component hypothesis, it does suggest that thermodynamic constraints were the most important drivers of ecological selection. We conceptualize thermodynamic constraints as arising from both the thermodynamic properties of the organic carbon pool (electron donors) and the availability of different terminal electron acceptors that differ in thermodynamic favorability (Boye *et al.* 2017). Direct and indirect relationships in the SEM model suggest specific linkages between electron donors and acceptors, that may be mediated by thermodynamic mechanisms. We note that inferences in the following discussion should be interpreted as speculative given that controlled experiments were not conducted, and that alternative interpretations of these relationships may be possible.

With respect to the organic carbon pool, we suggest that total carbon content accrues in the form of less favorable organic matter, as stocks of more favorable organic compounds are preferentially depleted (Graham *et al.* 2017); in turn, a relationship

emerges wherein soil carbon content increases with increasing values of $\Delta G^{\circ}_{Cox}(H_2O)$ (higher values indicate lower favorability (LaRowe and Van Cappellen 2011)). This is in agreement with a recent analysis of the permafrost microbiome along a chronosequence, which showed community adaptations related to decreasing availability of labile substrates (Mackelprang *et al.* 2017). These adaptations were associated with the enrichment in hydrocarbon degradation genes enabling catabolism of less labile carbon substrates in older permafrost (Mackelprang *et al.* 2017), consistent with our results suggesting that organic carbon in permafrost is non-randomly degraded with respect to its lability and thermodynamic favorability. The positive association between total carbon and nitrate further supports our interpretation of higher carbon content resulting from accumulation of organic molecules that are less thermodynamically favorable for microbially-driven organic carbon oxidation. In this case, higher total carbon reflects less thermodynamically favorable carbon, which would result in lower rates of nitrate reduction that depend on the oxidation of organic carbon, and thus a positive relationship between carbon and nitrate content.

Our data additionally suggest that organic acids, as products of anaerobic fermentation, support nitrate and Fe(III) reduction, and thus deplete the pools of these terminal electron acceptors. In this case, organic acid content is expected to be negatively associated with nitrate and positively associated with Fe(II), as observed. The reason for the positive relationship with Fe(II) is that Fe(III) (the terminal electron acceptor) is being depleted in the process of producing Fe(II). We thus infer that a positive relationship between organic acid content and Fe(II) indicates a negative relationship between organic acid content and bioavailable Fe(III). Additionally, we observed a negative relationship between nitrate and Fe(II), which likely arises because Fe(III) reduction is less thermodynamically favorable than nitrate reduction, and biological production of Fe(II) is thus only expected to be a dominant process in the absence of available nitrate.

CONCLUSIONS

Our findings support the hypothesis that permafrost communities are primarily shaped by the physical constraints imposed by prolonged freezing, and exhibit spatial distributions that reflect dispersal limitation and selective pressures associated with the physical constraints of the permafrost environment. We found that microbial distributions in permafrost are driven primarily by dispersal limitation and deterministic selection arising from thermodynamic constraints of the permafrost environment. This contrasts with non-permafrost soil communities, which appear to be driven primarily by soil pH (Fierer and Jackson 2006; Lauber *et al.* 2009; Chu *et al.* 2010). These findings suggest the need for different mechanistic models predicting microbial community characteristics in permafrost and non-permafrost soils, given the different processes governing these systems. Our findings suggest that predictive models of permafrost community composition will need to account for organic carbon thermodynamics, organic acid concentrations and redox conditions, which may be informed by knowledge of landscape history. However, efforts to accurately predict community composition at the landscape-scale based solely on environmental characteristics may be limited due to the strong influence of dispersal limitation.

Our findings additionally support the idea that changes in permafrost microbial community structure and function are likely to be drastic in response to thaw (Mackelprang *et al.* 2011), as hydrologic changes mobilize organisms and nutrients, thereby relieving the primary constraints on communities. Community responses to change are also likely to be highly varied across landscapes, given the environmental and microbiological heterogeneity of permafrost-affected soils. Identifying how pre-thaw environmental and community characteristics influence post-thaw responses will be essential for accurately predicting ecosystem level responses to environmental change.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGMENTS

We thank Caroline Anderson and Alex Crump for their assistance with sample collection and field work. We thank Tom Wietsma for conducting carbon and nitrogen analyses and Tom Resch for conducting metals and anion analyses.

FUNDING

This research was supported by the Microbiomes in Transition Initiative at Pacific Northwest National Lab (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy, and was conducted under the Laboratory Directed Research and Development Program at PNNL. A portion of this research was conducted using PNNL Institutional Computing resources and at the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility in Richland, WA.

Conflict of interest. None declared.

REFERENCES

Aronesty E. Comparison of sequencing utility programs. *Open Bioinforma J* 2013;**7**:1–8.

Austin AT, Yahdjian L, Stark JM *et al.* Water pulses and biogeochemical cycles in arid and semiarid ecosystems. *Oecologia* 2004;**141**:221–35.

Bakermans C, Skidmore ML, Douglas S *et al.* Molecular characterization of bacteria from permafrost of the Taylor Valley, Antarctica. *FEMS Microbiol Ecol* 2014;**2**:331–46.

Barberán A, Ladau J, Leff JW *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci USA* 2015;**112**:5756–61.

Bergmann GT, Bates ST, Eilers KG *et al.* The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* 2011;**43**:1450–55.

Bockheim JG. Permafrost distribution in the southern circumpolar region and its relation to the environment: a review and recommendations for further research. *Permafrost Periglac* 1995;**6**:27–45.

Borcard D, Gillet F, Legendre P. *Numerical Ecology with R*. New York: Springer-Verlag, 2011.

Borcard D, Legendre P, Avois-Jacquet C *et al.* Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 2004;**85**:1826–32.

Borcard D, Legendre P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* 2002;**153**:51–68.

Boye K, Noel V, Tfaily MM *et al.* Thermodynamically controlled preservation of organic carbon in floodplains. *Nat Geosci* 2017;**10**:415–19.

Brockett BF, Prescott CE, Grayston SJ. Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. *Soil Biol Biochem* 2012;**44**:9–20.

Brouchkov A, Kabilov M, Filippova S *et al.* Bacterial community in ancient permafrost alluvium at Mammoth Mountain (Eastern Siberia). *Gene* 2017;**636**:48–53.

Burn CR, Nelson FE. Comment on "A projection of severe near-surface permafrost degradation during the 21st century" by David M. Lawrence and Andrew G. Slater. *Geophys Res Lett* 2006;**33**:L21503.

Burrows SN, Gower ST, Clayton MK *et al.* Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems* 2002;**5**:0667–79.

Bushnell B. *BBMap: a fast, accurate, splice-aware aligner*. 2014. https://sourceforge.net/projects/bbmap.

Caporaso JG, Lauber CL, Walters WA *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;**6**:1621–24.

Chu H, Fierer N, Lauber CL *et al.* Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* 2010;**12**:2998–3006.

Cleveland CC, Nemergut DR, Schmidt SK *et al.* Increases in soil respiration following labile carbon additions linked to rapid shifts in soil microbial community composition. *Biogeochemistry* 2007;**82**:229–40.

Coolen MJ, van de Giessen J, Zhu EY *et al.* Bioavailability of soil organic matter and microbial community dynamics upon permafrost thaw. *Environ Microbiol* 2011;**13**:2299–314.

Coolen MJL, Orsi WD. The transcriptional response of microbial communities in thawing Alaskan permafrost soils. *Front Microbiol* 2015;**6**:197.

Cruz-Martínez K, Rosling A, Zhang Y *et al.* Effect of rainfall-induced soil geochemistry dynamics on grassland soil microbial communities. *Appl Environ Microbiol* 2012;**78**:7587–95.

Dini-Andreote F, Stegen JC, van Elsas JD *et al.* Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci USA* 2015;**112**:E1326–E32.

Duddleston KN, Kinney MA, Kiene RP *et al.* Anaerobic microbial biogeochemistry in a northern bog: acetate as a dominant metabolic end product. *Global Biogeochem Cy* 2002;**16**:11–1-11-9.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–61.

Ernakovich JG, Wallenstein MD, Calderón FJ. Chemical indicators of cryoturbation and microbial processing throughout an Alaskan permafrost soil depth profile. *Soil Sci Soc Am J* 2015;**79**:783–93.

Fierer N, Bradford MA, Jackson RB. Toward an ecological classification of soil bacteria. *Ecology* 2007;**88**:1354–64.

Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Nat Acad Sci USA* 2006;**103**:626–31.

Fox J, Zhenghua N, Byrnes J. *SEM: Structural Equation Models package version 3.1-9*. 2017. https://CRAN.R-project.org/package=sem.

Goordial J, Davila A, Lacelle D *et al.* Nearing the cold-arid limits of microbial life in permafrost of an upper dry valley, Antarctica. *ISME J* 2016;**10**:1613–24.

Graham EB, Tfaily MM, Crump AR *et al.* Carbon inputs from riparian vegetation limit oxidation of physically bound organic carbon via biochemical and thermodynamic processes. *J Geophys Res Biogeosci* 2017;**122**:3188–3205.

Griffiths RI, Thomson BC, James P *et al.* The bacterial biogeography of British soils. *Environ Microbiol* 2011;**13**:1642–54.

Hartmann A, Schmid M, Tuinen Dv *et al.* Plant-driven selection of microbes. *Plant Soil* 2009;**321**:235–57.

Hiemstra PH, Pebesma EJ, Twenhöfel CJW *et al.* Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput Geosci* 2009;**35**:1711–21.

Hines ME, Duddleston KN, Kiene RP. Carbon flow to acetate and C1 compounds in northern wetlands. *Geophys Res Lett* 2001;**28**:4251–54.

Hultman J, Waldrop MP, Mackelprang R *et al.* Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 2015;**521**:208–12.

Jansson JK, Tas N. The microbial ecology of permafrost. *Nat Rev Microbiol* 2014;**12**:414–25.

Jones RT, Robeson MS, Lauber CL *et al.* A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* 2009;**3**:442–453.

Kujawinski EB, Behn MD. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal Chem* 2006;**78**:4363–73.

LaRowe DE, Van Cappellen P. Degradation of natural organic matter: a thermodynamic analysis. *Geochim Cosmochim Acta* 2011;**75**:2030–42.

Lauber CL, Hamady M, Knight R *et al.* Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* 2009;**75**:5111–20.

Lee CK, Barbier BA, Bottos EM *et al.* The inter-valley soil comparative survey: the ecology of Dry Valley edaphic microbial communities. *ISME J* 2012;**6**:1046–57.

Lee H, Schuur EA, Inglett KS *et al.* The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. *Glob Change Biol* 2012;**18**:515–27.

Lipson DA, Jha M, Raab TK *et al.* Reduction of iron (III) and humic substances plays a major role in anaerobic respiration in an Arctic peat soil. *J Geophys Res Biogeosci* 2010;**115**:G00I06.

Lipson DA, Raab TK, Goria D *et al.* The contribution of Fe(III) and humic acid reduction to ecosystem respiration in drained thaw lake basins of the Arctic Coastal Plain. *Glob Biogeochem Cy* 2013;**27**:399–409.

Lumley T. *Leaps: Regression Subset Selection version 3*. 2017. https://CRAN.R-project.org/package=leaps .

Mackelprang R, Burkert A, Haw M *et al.* Microbial survival strategies in ancient permafrost: insights from metagenomics. *ISME J* 2017;**11**:2305–2318.

Mackelprang R, Saleska SR, Jacobsen CS *et al.* Permafrost meta-omics and climate change. *Annu Rev Earth Planet Sci* 2016;**44**:439–62.

Mackelprang R, Waldrop MP, DeAngelis KM *et al.* Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 2011;**480**:368–71.

McCalley C, Woodcroft B, Hodgkins S *et al.* Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* 2014;**514**:478–81.

McMahon PB, Chapelle FH. Microbial production of organic acids in aquitard sediments and its role in aquifer geochemistry. *Nature* 1991;**349**:233–35.

Miller KE, Lai C-T, Friedman ES *et al.* Methane suppression by iron and humic acids in soils of the Arctic Coastal Plain. *Soil Biol Biochem* 2015;**83**:176–83.

Minor EC, Steinbring CJ, Longnecker K *et al.* Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Org Geochem* 2012;**43**:1–11.

Mitzscherling J, Wingel M, Winterfeld M *et al.* The development of permafrost bacterial communities under submarine conditions. *J Geophys Res Biogeosci* 2017;**122**:1689–1704.

Mondav R, Woodcroft BJ, Kim E-H *et al.* Discovery of a novel methanogen prevalent in thawing permafrost. *Nat Commun* 2014;**5**:3212.

Oksanen J, Blanchet FG, Friendly M *et al. Vegan: community ecology package version 2.4-4*. 2017. https://CRAN.R-project.org/package=vegan

Park J-H, Matzner E. Controls on the release of dissolved organic carbon and nitrogen from a deciduous forest floor investigated by manipulations of aboveground litter inputs and water flux. *Biogeochemistry* 2003;**66**:265–86.

Ping CL, Jastrow JD, Jorgenson MT *et al.* Permafrost soils and carbon cycling. *Soil* 2015;**1**:147–71.

Ramirez KS, Craine JM, Fierer N. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Glob Change Biol* 2012;**18**:1918–27.

Rivkina E, Petrovskaya L, Vishnivetskaya T *et al.* Metagenomic analyses of the late Pleistocene permafrost-additional tools for reconstruction of environmental conditions. *Biogeosciences* 2016;**13**:2207–2219.

RStudioTeam. *RStudio: Integrated Development for R*. Boston: RStudio, Inc., 2015.

Schuur EAG, Bockheim J, Canadell JG *et al.* Vulnerability of permafrost carbon to climate change: implications for the global carbon cycle. *BioScience* 2008;**58**:701–14.

Shur Y, Hinkel KM, Nelson FE. The transient layer: implications for geocryology and climate-change science. *Permafrost Periglac* 2005;**16**:5–17.

Stegen JC, Lin X, Fredrickson JK *et al.* Estimating and mapping ecological processes influencing microbial community assembly. *Front Microbiol* 2015;**6**:370.

Stegen JC, Lin X, Fredrickson JK *et al.* Quantifying community assembly processes and identifying features that impose them. *ISME J* 2013;**7**:2069–79.

Steven B, Briggs G, McKay CP *et al.* Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol Ecol* 2007;**59**:513–23.

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

Stookey LL. Ferrozine — a new spectrophotometric reagent for iron. *Anal Chem* 1970;**42**:779–81.

Tarnocai C, Canadell JG, Schuur EAG *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob Biogeochem Cy* 2009;**23**:GB2023.

Tas N, Prestat E, McFarland JW *et al.* Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J* 2014;**8**:1904–19.

Tfaily MM, Chu RK, Tolić N *et al.* Advanced solvent based methods for molecular characterization of soil organic matter by high-resolution mass spectrometry. *Anal Chem* 2015;**87**:5206–15.

Tfaily MM, Chu RK, Toyoda J *et al.* Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal Chim Acta* 2017;**972**:54–61.

Treves DS, Xia B, Zhou J *et al.* A two-species test of the hypothesis that spatial isolation influences microbial diversity in soil. *Microbial Ecol* 2003;**45**:20–28.

Trivedi P, Anderson IC, Singh BK. Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. *Trends Microbiol* 2013;**21**:641–51.

Tuorto SJ, Darias P, McGuinness LR *et al.* Bacterial genome replication at subzero temperatures in permafrost. *ISME J* 2014;**8**:139.

Van Everdingen RO. *Multi-language Glossary of Permafrost and Related Ground-ice Terms in Chinese, English, French, German, Icelandic, Italian, Norwegian, Polish, Romanian, Russian, Spanish, and Swedish*. Calgary: International Permafrost Association, Terminology Working Group, 1998.

Vaughan DG, Comiso JC, Allison I *et al.* Observations: cryosphere. In: Stocker TF, Qin D, Plattner G-K *et al.* (eds). *Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, 2013,317–82.

Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol* 2010;**85**:183–206.

Waldrop MP, Wickland KP, White III R *et al.* Molecular investigations into a globally important carbon pool: permafrost-protected carbon in Alaskan soils. *Glob Change Biol* 2010;**16**:2543–54.

Wieder WR, Allison SD, Davidson EA *et al.* Explicitly representing soil microbial processes in Earth system models. *Glob Biogeochem Cy* 2015;**29**:1782–800.

Yergeau E, Hogues H, Whyte LG *et al.* The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarrays. *ISME J* 2010;**4**:1206–14.

Zachara JM, Moran JJ, Resch CT *et al.* Geo- and biogeochemical processes in a heliothermal hypersaline lake. *Geochim Cosmochim Acta* 2016;**181**:144–163.

Zhang T, Barry RG, Knowles K *et al.* Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere. *Polar Geogr* 2008;**31**:47–68.

Zhao C, Miao Y, Yu C *et al.* Soil microbial community composition and respiration along an experimental precipitation gradient in a semiarid steppe. *Sci Rep* 2016;**6**:24317.