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, Richland, WA, USA

[2] Department of Biological Sciences, Thompson Rivers University, Kamloops, BC, Canada

[3] Computational Biology, Pacific Northwest National Laboratory, Richland, WA, USA

[4] National Security Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

[5] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

***Corresponding Author:**

Eric M. Bottos

Email: ebottos@tru.ca

Thompson Rivers University

Box 3010, 900 McGill Rd

Kamloops, BC, V2C 0C8

Abstract

Understanding drivers of permafrost microbial community composition is critical for understanding permafrost microbiology and predicting ecosystem responses to thaw, however studies describing ecological controls on these communities are lacking. We hypothesize that permafrost communities are uniquely shaped by constraints imposed by prolonged freezing, and decoupled from factors that influence non-permafrost soil communities. To test this hypothesis, 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 relative 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 reveal that the processes shaping microbial communities in permafrost are distinct from those in non-permafrost soils, as the stability of the permafrost environment imposes dispersal and thermodynamic constraints on permafrost communities. Models of permafrost community composition will need to

47 account for these unique drivers in order to predict community characteristics
48 across permafrost landscapes, and in efforts to understand how pre-thaw conditions
49 will influence post-thaw ecological and biogeochemical processes.
50

Introduction

Permafrost is defined as ground that has remained below 0 °C for two or more consecutive years [1]. Because this definition is solely based on a condition of 'ground climate' [2], 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 [3], spanning much of the terrestrial Arctic and subarctic [4], ice-free areas of Antarctica [5], and high-elevation regions in both the northern and southern hemispheres [4,6]. Collectively, these soils represent an important microbial ecosystem [7] and a globally significant pool of sequestered carbon [3,8], which is being mobilized as climate warming increases permafrost thaw [9]. 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.

While both environmental conditions and microbial community composition of permafrost-affected soils are known to be highly variable, the degree to which variation in community composition is linked to physicochemical conditions of the soil is not well understood [10]. In many non-permafrost soils, microbial community composition is shaped by physicochemical conditions, including pH [11-13], nutrient content [14,15], and soil moisture [16-18]. Given that permafrost can support active microbial communities [19,20], it is reasonable to assume that similar factors may be

74 important in structuring the permafrost microbiome. Alternatively, microbial
75 community composition in these environments may be decoupled from
76 physicochemical conditions that are found to be important in non-permafrost soils,
77 and may instead be similarly shaped by the shared constraints imposed by
78 prolonged freezing. Understanding how microbial communities are shaped by
79 environmental conditions represents an important knowledge gap in permafrost
80 microbiology.

81 While resolving drivers of community composition in permafrost
82 environments will improve fundamental understanding of the microbiology of these
83 extreme ecosystems, there is also practical importance in resolving how pre-thaw
84 conditions may be used as predictors of system level response to thaw. Earth system
85 models that integrate aspects of microbial community composition and function are
86 gaining support to improve understanding of terrestrial carbon cycling and
87 predictions about the fate of soil carbon in response to environmental change [21,22].
88 However, permafrost environments bring a high level of complexity that is difficult
89 to generalize in current models, because soil type, soil conditions, and carbon
90 composition may all have important impacts on post thaw dynamics and carbon
91 transformation [23-26]. Additionally, the composition of pre-thaw communities may be
92 a strong determinant of post-thaw processes, as permafrost microbial communities
93 are expected to respond rapidly to thaw [27,28], and the abundance of particular taxa
94 and functional genes can be important predictors of process rates, such as
95 methanogenesis [26,29,30] and iron reduction [31]. These findings underscore the
96 importance of integrating knowledge of the physical environment, the chemical

97 nature of the organic matter pool, and the structure and function of permafrost
98 microbial communities to accurately predict rates of carbon metabolism in these
99 systems. Spatially explicit studies capturing measures of soil heterogeneity are,
100 therefore, necessary to inform models aimed at predicting microbial community
101 responses to permafrost thaw and carbon fate in these environments.

102 The purpose of this work was to resolve the factors and processes that
103 govern microbial community structure in permafrost-affected soils. We hypothesize
104 that the factors and processes shaping permafrost microbial communities differ
105 from those shaping non-permafrost soil communities, and reflect the unique
106 constraints of the permafrost environment. We characterized patterns of microbial
107 community composition along landscape gradients in the boreal forest ecosystem of
108 the Caribou Poker Creek Research Watershed (CPCRW) near Fairbanks, AK. We
109 examined the influence of dispersal and selection on patterns of community
110 composition and evaluated the importance of permafrost physicochemical
111 conditions, including soil organic matter composition and thermodynamic
112 properties, as deterministic factors. As the first landscape-scale survey linking
113 permafrost community composition to environmental variability, this work
114 provides mechanistic understanding of the controls on permafrost communities.
115 This understanding can, in turn, inform models aimed at predicting permafrost
116 microbial community characteristics and responses to thaw.

117

118 **Materials and Methods**

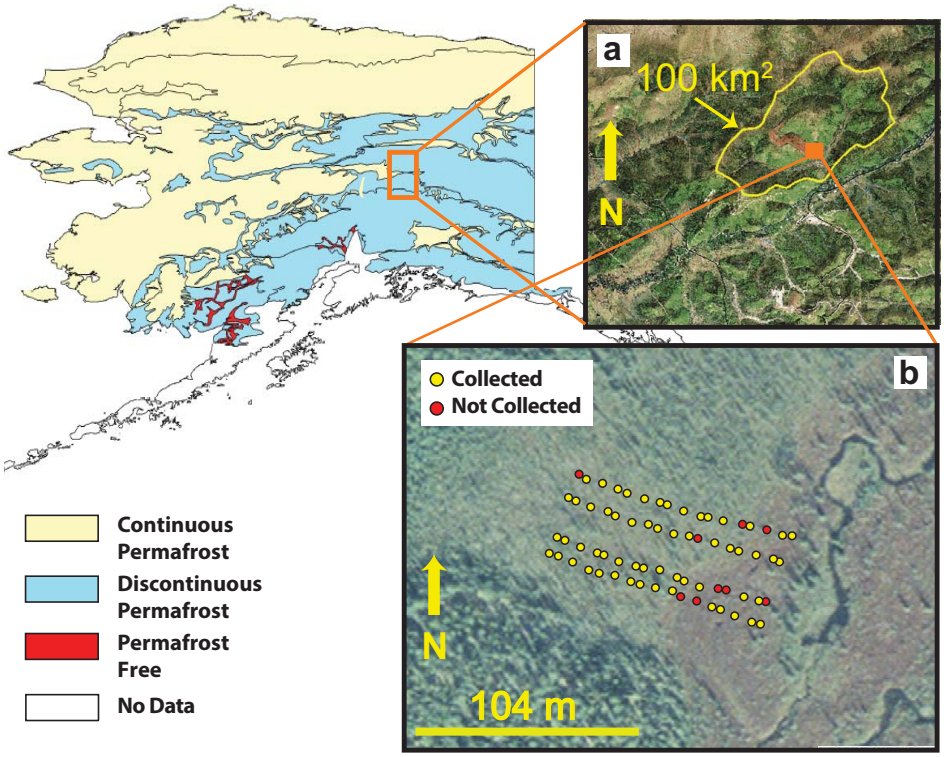
119 *Sample collection and processing*

Samples were collected along a hydrologic gradient in the Caribou Poker Creek Research Watershed (CPCRW). CPCRW is a long-term ecological research (LTER) 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. To efficiently capture spatial variation at the landscape scale, we used a cyclic sampling design, as opposed to regular grid spacing [32]. 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 (Figure 1).

At each sampling position, permafrost cores were collected using a SIPRE coring auger (John's Machine Shop, Fairbanks, AK). 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 3-12 cm of each core was removed using an ethanol sterilized chisel and the next 4-10 cm section of each core was taken for analysis. 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.

143 Samples to be stored anaerobically were immediately purged with nitrogen gas in
144 60 ml serum bottles sealed with butyl rubber stoppers, and all partitioned material
145 was stored at -20 °C until analysis. A total of 59 samples were included in the final
146 analyses, as cores could not be retrieved from some sample locations or were
147 compromised during sample processing (refer to Figure 1).



148

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

155

156

157 *Physicochemical analyses*

158 Soil water content was determined by drying 1-10 g of sample at 105 °C, and
159 measuring mass loss after 48 hours: sample masses were determined after 24 and
160 48 h to ensure samples reached a constant mass in consecutive measurements.
161 Water content was determined from the average of five replicate measurements per
162 sample.

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

167 Samples for metals and anion analyses were prepared from freeze-dried,
168 ground, and <2 mm sieved samples. For metals analysis, 1 g of sample was extracted
169 with 10 ml of 0.5 N HCl shaking at 200 rpm for 2 h at room temperature. Anion
170 extractions were completed as above, with 1 g of soil in 5-10 ml of deionized water.
171 Metals (Fe, Mn, Mg, Cu, P, S) and anion (Cl^-, SO_4^-, and NO_3^-) analyses were completed
172 as previously described (Zachara et al., 2009).

173 Iron(II) content was determined by ferrozine assay [33]. In an anaerobic
174 chamber (Coy Laboratory Products, Grass Valley, MI), 10 ml 0.5 N HCl was added to
175 1 g of permafrost sample, and the vial was sealed and vortexed. Samples were
176 extracted for 1 h and filtered through a 0.22 μm pore-size polyethersulfone (PES)
177 syringe filter. Extracts were diluted in 0.1 N HCl and 100 μl was added to 1 ml
178 ferrozine; after 5 min, the absorbance at 562 nm was measured on a Shimadzu

179 Biospec-1601 spectrophotometer. Iron(II) concentrations were determined from a
180 six-point standard curve ranging from 0 to 45 μM Iron(II). Samples were dried at 60
181 °C and weighed to determine the Iron(II) content by dry weight.

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

190 Soil texture was determined by measuring the gravel (> 2mm), sand (64 μm-
191 2 mm), and mud (silt and clay) (<64 μm) fractions of each sample. Briefly, 20 g of
192 soil was dried at 60 °C, and the total dry weight determined. Samples were dry
193 sieved through a 2 mm sieve, and the mass of the >2 mm fraction was used to
194 calculate the gravel fraction. The <2 mm fraction was wet sieved through a 64 μm
195 sieve, the fraction retained was dried and used to calculate the sand fraction, while
196 the <64 μm fraction was dried and used to calculate the mud fraction.

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

200

201 *Organic matter composition determination by FT-ICR-MS*

Organic matter was extracted from bulk soil sequentially with water, methanol and chloroform as described previously [34,35]. Briefly, organic matter extracts were prepared by adding 1 ml of solvent to 100 mg lyophilized and ground bulk soil and shaking for two hours 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 (EMSL) 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 s 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 (CIA) [36], and modified as previously described [37]. 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 ($\Delta G°_{Cox}$) of each compound was derived based on the nominal oxidation state of carbon (NOSC) as previously described [38]. The average $\Delta G°_{Cox}$ of the carbon pool was determined for each sample extraction: the median values were used for the methanol ($\Delta G°_{Cox}(MeOH)$) and chloroform ($\Delta G°_{Cox}(CHCl_3)$)

247 extracts due to highly skewed distributions, while the $\Delta G°_{Cox}$ was normally
248 distributed for water extracts ($\Delta G°_{Cox}(H_2O)$) such that the mean values were used.
249
250 *Microbial community analyses*
251 Total community DNA was extracted from 0.25 g of each sample using the
252 MoBio Power Soil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA), according
253 to manufacturer's instructions. Additional cleanup and concentration of DNA
254 extracts was completed using the Zymo ZR-96 Genomic DNA Clean and
255 Concentrator-5 kit (Zymo Research Corporation, Irvine, CA). PCR amplification of
256 the V4 region of the 16S rRNA gene was performed as previously described [39], with
257 the exception that the twelve-base barcode sequence was included in the forward
258 primer. Amplicons were sequenced on an Illumina MiSeq using the 500 cycle Miseq
259 Reagent Kit v2 (Illumina Inc., San Diego, CA), according to manufacturer's
260 instructions.
261 Raw sequence reads were demultiplexed using EA-Utils [40] not allowing any
262 mismatches in the barcode sequence. Reads were quality filtered with BBDuk2 [41] to
263 remove adapter sequences and PhiX with matching kmer length of 31 bp at a
264 hamming distance of 1. Reads shorter than 51 bp were discarded. Reads were
265 merged using USEARCH [42] with a minimum length threshold of 175 bp and
266 maximum error rate of 1 %. Sequences were de-replicated and clustered using the
267 distance-based, greedy clustering method of USEARCH at 97 % pairwise sequence
268 identity among operational taxonomic unit (OTU) member sequences. Taxonomy
269 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 17899 sequences and OTUs that could not be assigned at the kingdom level were removed.

Statistical analysis

The environmental variables consisted of all physicochemical variables and the average $\Delta G°_{Cox}$ for each FTICR extraction (mean for $\Delta G°_{Cox}(H_2O)$ and median for $\Delta G°_{Cox}(MeOH)$ and $\Delta G°_{Cox}(CHCl_3)$. 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, $\Delta G°_{Cox}(H_2O)$, $\Delta G°_{Cox}(CHCl_3)$, and $\Delta G°_{Cox}(MeOH)$ were not transformed.

Principal components analysis (PCA) was used to assess variation in environmental variables across the landscape using the *princomp* function in R [43]. 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.

293 Analyses of community diversity and composition were completed using the
294 '*vegan'* package [44] in R. Shannon diversity estimates were completed based on the
295 resampled OTU counts. OTU abundances were Hellinger transformed prior to all
296 other compositional analyses. Non-metric multidimensional scaling was used to
297 examine the community variation between samples based on Bray Curtis
298 dissimilarity, and environmental variables were fit as vectors in the final two-
299 dimensional ordination to evaluate relationships between community and
300 environmental variation.

301 Spatial analyses were completed in R. Kriging was used to interpolate and
302 visualize spatial trends in both the environmental and biological data, using the
303 autokrig function of the *'automap'* package [45]. Principal coordinates of neighbor
304 matrices (PCNM) was used to create orthogonal spatial variables based on sample
305 site locations [46,47]. PCNMs were calculated as previously described [48]. PCNM axes
306 were used as explanatory variables in downstream analyses to examine the
307 importance of spatial filters on community composition.

308 A redundancy analysis (RDA) model was used to relate community
309 composition to environmental and spatial variation using the *vegan* package [44] in R.
310 Due to collinearity between several environmental variables, PC scores extracted
311 from the environmental PCA were used to represent environmental variables in the
312 model. Forward stepwise model building based on adjusted R^2 was carried out using
313 all 23 PCs and all 15 positive PCNMs. The importance of each variable added to the
314 model was assessed using variance partitioning based on RDA.

315 Null modeling was used to estimate the influence of ecological processes on
316 community composition, as described previously [49,50]. The influence of selection was
317 estimated by evaluating the difference between the observed between-community
318 mean-nearest-taxon distance (βMNTD) and the mean of the null distribution in units
319 of standard deviation. Significant deviations from the null distribution were
320 evaluated using the β–nearest taxon index (βNTI) and the signal for selection was
321 expressed as the proportion of comparisons for which βNTI>2 or βNTI<-2,
322 representing signals for variable selection and homogenous selection, respectively.
323 Comparisons falling within the null distribution (2>βNTI>-2) represent
324 compositional differences that do not arise from selection, and are instead
325 attributable to dispersal limitation, homogenizing dispersal, or processes
326 undominated by dispersal or selection. To assess the relative influence of these
327 processes, a Raup-Crick metric incorporating species relative abundance (RC_{bray})
328 was used to compare the observed and expected species turnover between
329 communities. Significant deviations from the null distribution indicating greater
330 than expected differences in community composition (2>βNTI>-2 and RC_{bray}>0.95)
331 were attributed to dispersal limitation, while those indicating less than expected
332 differences in community composition (2>βNTI>-2 and RC_{bray}<-0.95) were
333 attributed to homogenizing dispersal. Comparisons falling within the null
334 distribution of both metrics (2>βNTI >-2 and 0.95>RC_{bray}>-0.95) represent
335 differences in community composition that were not strongly governed by selection
336 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 [50]. Community-level estimates of total selection were then used as the dependent variable in an exhaustive model selection using Bayesian information criterion (BIC), performed in the '*leaps'* package [51] 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 [52] in R (Figure S1). 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.

Code availability

Custom computer code used in the current study is available from the corresponding author on reasonable request.

Data availability

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

and analyzed in the current study are available from the corresponding author on reasonable request.

Results

Environmental conditions and carbon composition

Permafrost characteristics were highly variable across the sampling area, and are summarized in Table S1. 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 (Table S2). 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 6.61 %, 8.93 %, and 4.96 % 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 $\Delta G°_{Cox}$ estimates from the FTICR profiles were tightly linked to the overall variation in FTICR compound classes for each extraction (Figure S2). The $\Delta G°_{Cox}$ 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 (Figure 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), $\Delta G°_{Cox}(H_2O)$ (-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).

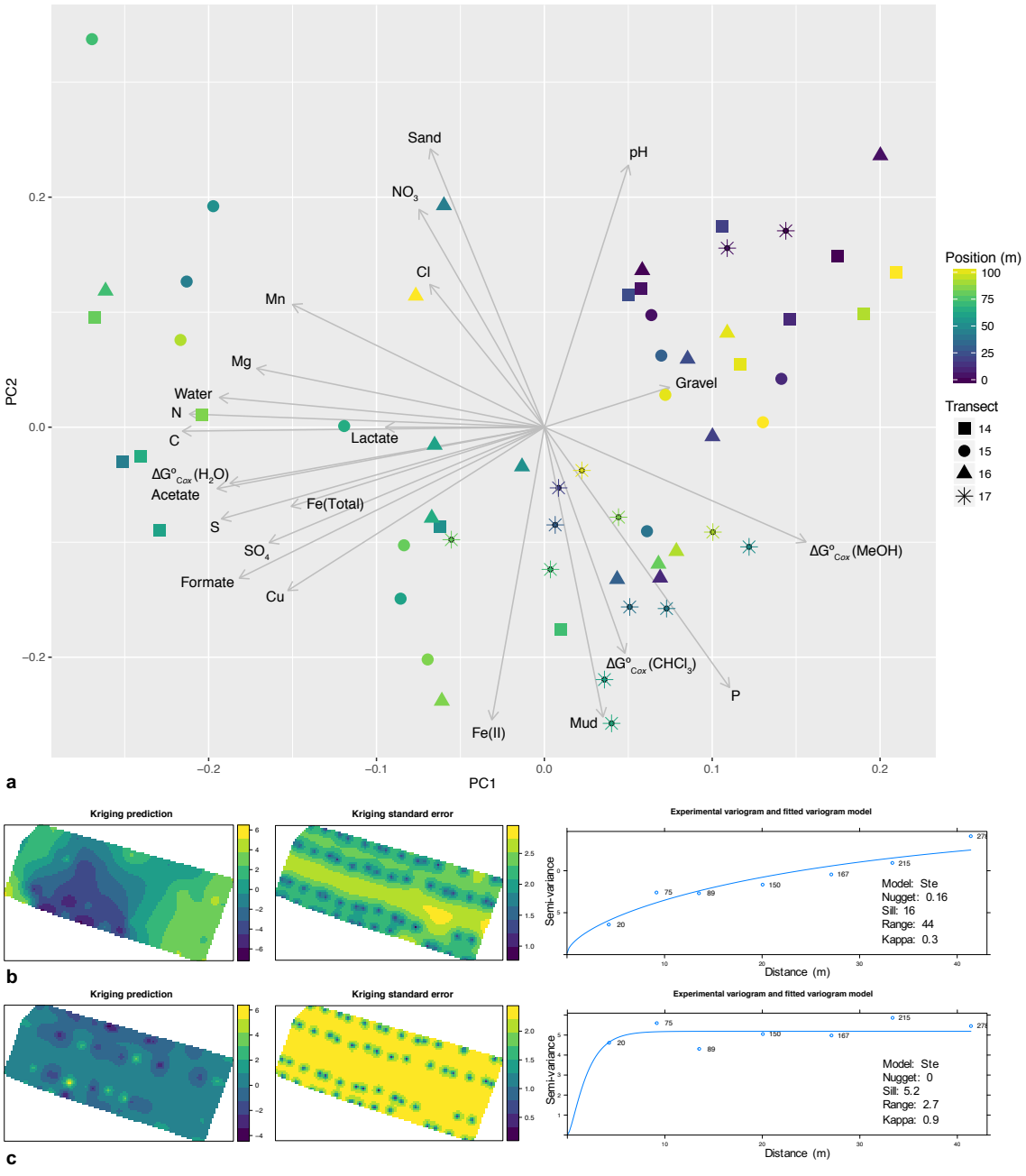


Figure 2: a) Principal components analysis (PCA) representing environmental variation between samples, and Kriging predictions of spatial patterns of environmental variation based on b) PC1 and c) PC2 scores across the sampling area (n=59).

404

405 *Microbial community composition*

406 Bacterial sequences grouped to a total of 45 phyla or candidate phyla, and 11
407 phyla were represented at >1 % of the total community (Figure S3). Based on the
408 average number of sequences in each sample grouping to bacterial phyla,
409 communities were dominated by Proteobacteria (23.9 %) (particularly Beta- (10.9
410 %), Alpha-(5.9 %), Delta- (5.3 %), and Gamma- (1.5 %) proteobacteria),
411 Acidobacteria (16.9 %), Verrucomicrobia (13.4 %), Actinobacteria (9.9 %),
412 Chloroflexi (9.8 %), Bacteroidetes (8.5 %), Gemmatimonadetes (5.0 %),
413 Planctomycetes (1.9 %), Nitrospirae (1.5 %), Parcubacteria (1.4 %), and Firmicutes
414 (1.1 %). Bacterial sequences grouping to other phyla and bacterial sequences that
415 could not be classified at the phylum level represented 4.8 % and 0.8 % of the total
416 sequences, respectively.

417 Approximately 1.2 % of sequences were classified as Archaeal, with 79 % of
418 these sequences grouping to the phylum Euryarchaeota. Sequences within the
419 Euryarchaeota grouped predominantly within the Methanomicrobia and
420 Methanobacteria.

421

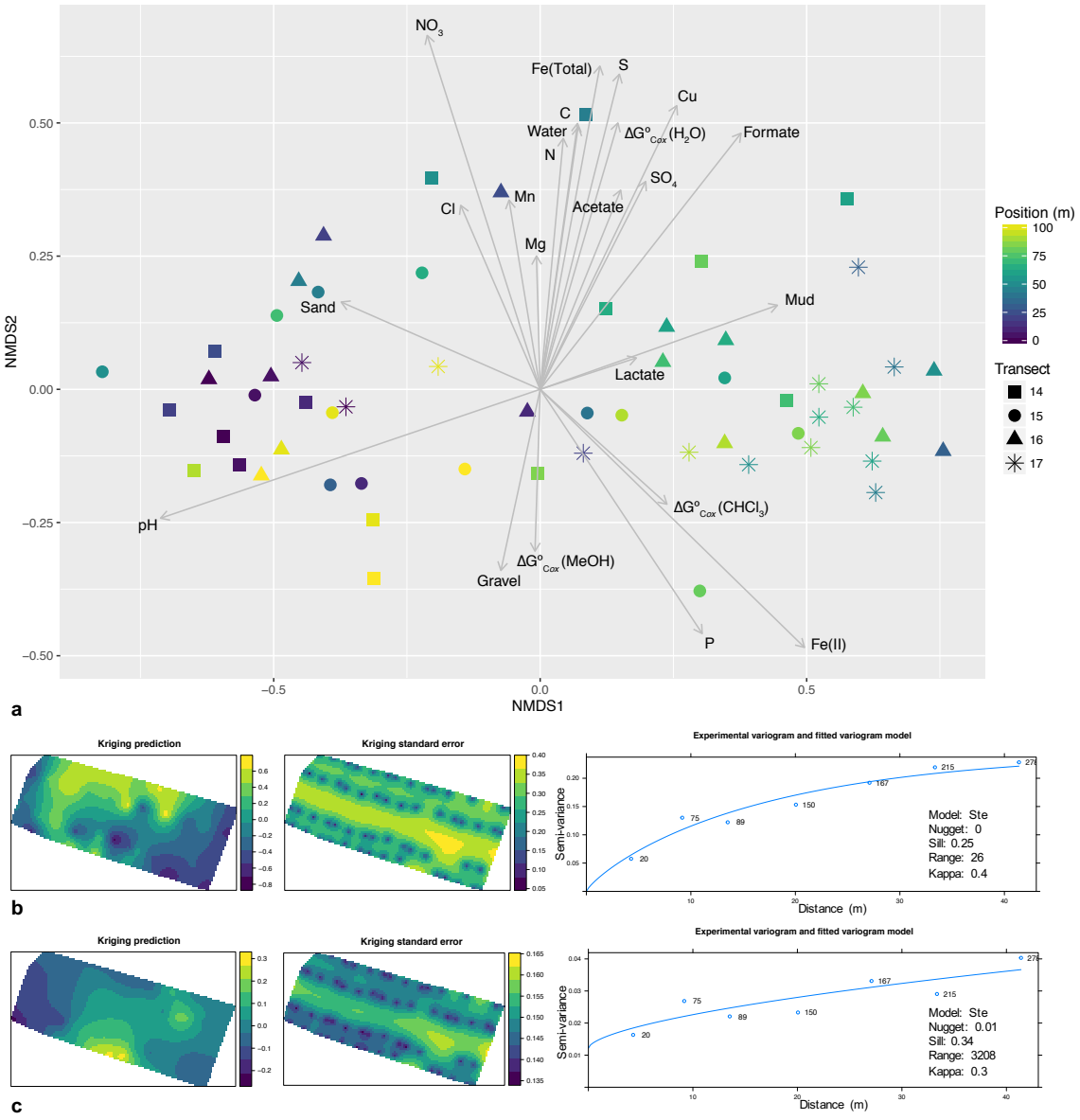
422 *Patterns of community composition*

423 Community composition showed non-random spatial structure (Figure 3),
424 and was explained by both environmental variables (PCs) and, to a lesser extent,
425 spatial variables (PCNMs). Stepwise model selection supported a model with fifteen
426 variables, which fit the data with an adjusted R^2=0.48; however, variance

427 partitioning showed that many of these variables contributed only incrementally to
428 improving model fit (Figure S4). A model incorporating the first two variables from
429 the selected model (PC2, PC1) fit the data with an adjusted R^2=0.34, and subsequent
430 addition of the remaining variables retained in the selected model improved the
431 adjusted R^2 by 0.02 (PCNM4) or less (all other variables) (Figure 4). We therefore
432 focused our interpretation on the model including PC2, PC1, and PCNM4. The
433 variable loadings on the PCs selected in the model indicated several environmental
434 factors were related to community composition (see Figure 2 for relationships
435 between environmental factors along PC1 and PC2).

436 Univariate regression of factors with the strongest loadings along PC1 and
437 PC2 showed that alpha diversity and the relative abundances of particular taxa were
438 significantly associated with one or more of these environmental variables (Table
439 S3).

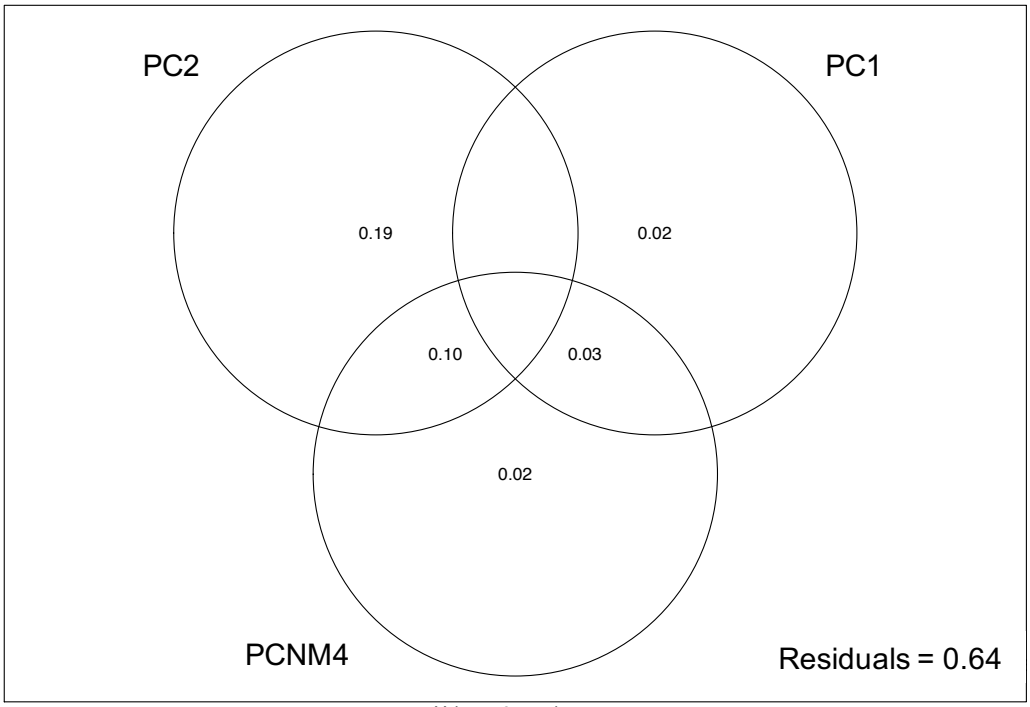
440



441

442 Figure 3: a) Non-metric multidimensional scaling (NMDS) plot representing the
443 Bray Curtis dissimilarity in microbial community composition between samples,
444 with environmental vectors overlaid, and Kriging predictions of spatial patterns of
445 community composition based on b) NMDS1 and c) NMDS2 scores across the
446 sampling area (n=59).

447



448

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

453

454

455 *Null model analyses*

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

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

470

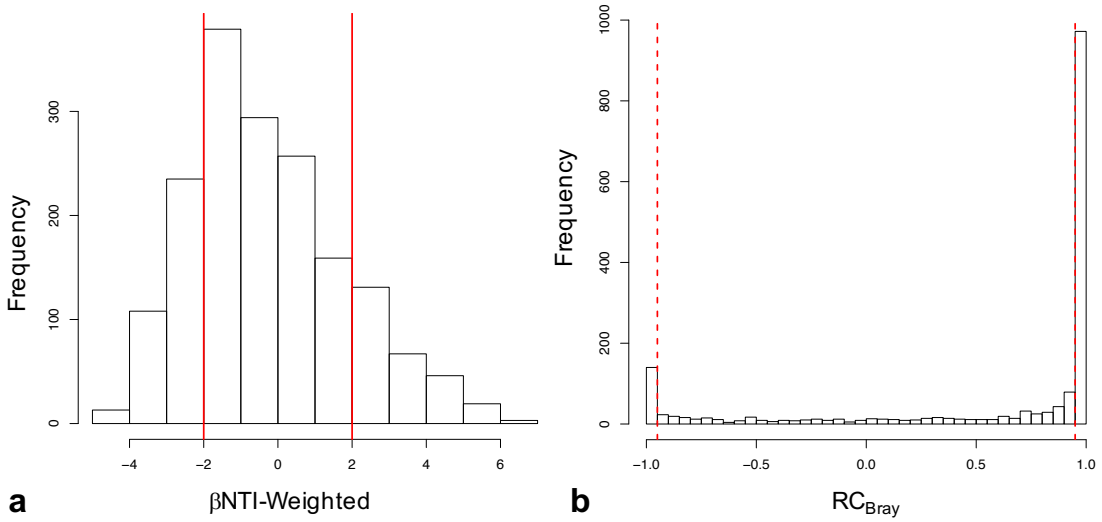


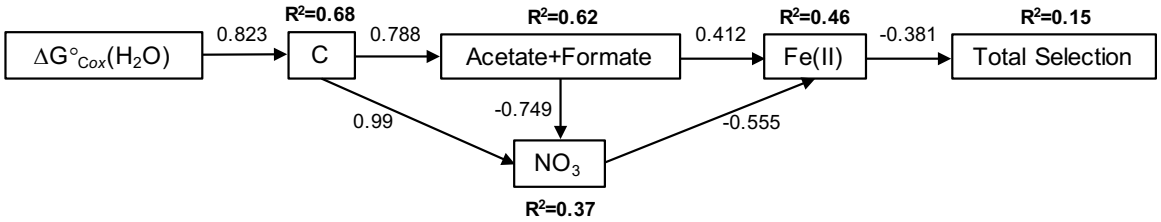
Figure 5: Histograms representing the observed distribution of comparisons based on a) βNTI and RC_{Bray}. Red lines represent the significance thresholds, whereby values outside their bounds are significantly different from the null distribution (n=59).

477

478 *Path analysis*

479 Given the relationship observed between Fe(II) and total selection, we
480 proposed a path model in which variables that reflect energetic constraints on
481 microbial activity may influence total selection indirectly, through relationships
482 mediated by Fe(II) content (See Figure S1). Our initial model was not consistent
483 with the data (X^2=37.2, d.f.=9 p=2.4x10^{-5}), and was revised to better reflect
484 relationships between the variables. All paths in the initial model were retained in
485 the final model, and modification indices supported the addition of a path from soil
486 carbon content to nitrate content. The final model did not differ significantly from
487 the data (X^2=10.3, d.f.=8, p=0.25) and explained 14.5 % of the variation in total
488 selection and between 37 and 68 % of the variation in other endogenous variables
489 (Figure 6). The direct effect of Fe(II) was the strongest total effect on total selection,
490 while organic acid content had the strongest indirect effect on total selection (Table
491 1).

492



493

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

499

500 **Table 1**: Standardized direct effects, indirect effects, and total effects of
501 environmental factors on total selection.

	Direct Effects	Indirect Effects	Total Effects
$\Delta G°_{Cox}(H_2O)$		-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

502

503

504 **Discussion**

505 Previous studies have demonstrated that permafrost soils contain diverse
506 and varied communities that are likely active *in situ*; however, studies of permafrost
507 microbiology typically suffer from low sample sizes, limiting the ability to examine
508 ecological relationships that may influence community structure and function. To
509 address this knowledge gap, we characterized patterns of environmental variation
510 and microbial community composition in a boreal forest ecosystem across
511 landscape gradients. By employing a well-replicated and geostatistically-informed
512 sampling design, we have provided the first characterization of ecological processes
513 driving landscape scale spatial structure of permafrost microbial community
514 composition. Through this work, we show that patterns of both environmental
515 characteristics and microbial community composition can be highly variable over
516 short distances and exhibit non-random spatial structure, with patterns of
517 community composition driven by deterministic and neutral processes that arise
518 primarily from the physical constraints of the permafrost environment.

519

520 *Spatial structure of permafrost physicochemistry*

521 The degree of heterogeneity in soil physicochemical and organic matter
522 characteristics observed over the study area was striking, likely reflecting spatially
523 structured variation in thaw history and organic matter deposition. We observed a
524 non-linear spatial trend in environmental variation, with samples at the extreme
525 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, $\Delta G°_{Cox}(H_2O)$, 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 [53]. Total carbon and nitrogen content, $\Delta G°_{Cox}(H_2O)$, 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 were 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 $\Delta G°_{Cox}$ compounds, such as lignin-like compounds, increasing the average $\Delta G°_{Cox}$ of the carbon pool.

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 $\Delta G°_{Cox}$. In sediments, a net accumulation of organic acids is observed when fermentation rates exceed respiration rates [54], and acetate and C_1 compounds are the dominant organic products of anaerobic metabolism in northern wetlands and bogs [55,56]. These products of anaerobic metabolism may accumulate in permafrost through equivalent processes.

549

550 *Microbial community composition and environmental correlates*

551 Community composition across our study site shared similarities with
552 permafrost communities reported previously from across the Arctic, suggesting a
553 core permafrost microbiome may be selected for by shared environmental
554 constraints across disparate locations. We observed high representation of
555 Acidobacteria and Proteobacteria, consistent with permafrost communities
556 reported from Sweden [30], and parts of Alaska [31], and high representation of
557 Chloroflexi, which has recently been reported in other Alaskan permafrost samples
558 [28,31]. Archaeal communities represented only a small percentage of the libraries and
559 were dominated by taxa grouping to methanogens in the phylum Euryarchaeota,
560 which is consistent with previous reports from across the Arctic [28,30,31,57]. In contrast
561 with previous studies, we saw high representation of Verrucomicrobia, which are
562 globally abundant in soils [58], but have not been previously reported as dominant
563 members of permafrost communities. Further comparisons of geographically
564 distinct permafrost communities will require an increased number of studies
565 employing well-replicated sampling designs and the adoption of standardized
566 analytical techniques within the field [7,10].

567 Permafrost communities across the study site were influenced by similar
568 drivers to non-permafrost soil communities, however several relationships were
569 indicative of the unique constraints of the permafrost environment. Diversity was
570 best described by a positive linear relationship with pH, which is consistent with
571 trends observed in non-permafrost soils [11-13]. The overall variation in community

572 composition showed a clear relationship with environmental variation (PCs),
573 although the particular environmental variables influencing community variation
574 were not clear from the RDA model selection. Despite a relatively broad range of pH
575 values across samples, no strong relationship between pH and community
576 composition was observed. The relative abundance of the dominant phyla also
577 varied significantly with numerous environmental variables, however these
578 relationships were atypical of trends observed in surveys of non-permafrost soils.
579 For example, at the phylum level, Acidobacteria are typically negatively associated
580 with pH, while Bacteroidetes and Actinobacteria typically have positive
581 relationships with pH [11]; however, we observed the opposite trends for both
582 Acidobacteria and Bacteroidetes and no trend for Actinobacteria with soil pH.
583 Selective constraints of the permafrost environment may limit the phylogenetic
584 breadth of these taxa, altering phylum level trends from those observed in other
585 soils. Additionally, other deterministic factors, such as soil redox conditions and soil
586 organic matter composition, which also showed strong univariate relationships with
587 the relative abundance of particular taxa, may be more important drivers of
588 community structure in permafrost-affected soils.
589
590 *Ecological processes influencing community composition*
591 We employed a null modelling approach to evaluate the degree to which
592 deterministic processes drive community variation and to resolve the variables
593 most likely to be causally influencing composition. Patterns of community
594 composition arise from a combination of deterministic and stochastic process [59] and

595 the relative importance of these processes vary between systems. Null modeling
596 provides a valuable tool to disentangle the influence of individual processes on
597 patterns of microbial distribution [50,60]. This approach has significant advantages
598 over the RDA models, which cannot estimate relative contributions of assembly
599 processes and did not reveal specific environmental variables that drive spatial
600 variation in community composition.

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

612 Given strong dispersal limitation, we expected that the spatial PCNM
613 variables would explain significant variation in community composition,
614 independent of environmental variation. This expectation was not met, however,
615 with PCNM axes explaining little variation in the RDA model. The lack of a strong
616 spatial signal in the RDA model indicates that the influence of spatial processes

617 manifest below the spatial resolution of our sampling, consistent with very
618 restricted movement of microorganisms through permafrost.

619 We found 37 % of the total community variation in permafrost community
620 composition was explained by selective processes, and that soil characteristics
621 associated with Fe(II) content are likely the most important environmental
622 variables deterministically influencing community composition. Soil Fe(II)
623 accumulates in anaerobic soils through the reduction of Fe(III), and iron reduction
624 can contribute substantially to respiration in Arctic soils [61,62]. A recent multi-omic
625 analysis of Alaskan permafrost reported high representation of proteins annotated
626 to iron-reducing taxa and the expression of genes annotated as cytochromes central
627 to iron-reduction, suggesting iron-reducing taxa were likely active *in situ* [31].
628 Importantly, Fe(III) reduction competes with other anaerobic processes, and
629 suppresses less thermodynamically favorable methanogenic pathways [63]. The
630 relationship between Fe(II) and total selection indicates that soil redox conditions
631 and thermodynamic constraints on microbial metabolism are likely to be the
632 primary selection pressures that deterministically govern community composition.

633 These findings suggest that the stability of the permafrost environment
634 strongly influences community structure and function, directly by restricting
635 community mixing and indirectly by influencing the selective landscape, as electron
636 donors and acceptors are depleted and infrequently replenished. This contrasts
637 strongly with non-permafrost soils, in which communities are presumed to be well-
638 dispersed through aeolian [64] and hydrologic process [65], nutrient fluxes are dynamic
639 [66,67], and communities are thought to be shaped predominantly by selection [68,69].

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

The thermodynamic constraints driving selection arise from the composition of both the organic matter and terminal electron acceptor pools, as outlined in our final SEM. We suggest that total carbon content accrues in the form of less favorable organic matter, as stocks of more favorable organic compounds are depleted; in turn, a relationship emerges wherein soil carbon content increases with $\Delta G°_{Cox}$ (higher values indicate lower favorability [38]). Further, organic acids are expected to accumulate in these soils through anaerobic metabolism, as labile carbon is fermented. We suggest that these organic acids support nitrate and Fe(III) reduction, such that organic acid content is negatively associated with nitrate, and positively associated with Fe(II). Additionally, a negative relationship between nitrate and Fe(II) likely arises because Fe(III) reduction is less energetically favorable than nitrate reduction. Modification indices supported a positive association between total soil carbon and nitrate content in the final model: the positive association between total carbon and nitrate is consistent with 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 favorable carbon, which would result in lower rates of nitrate reduction that depend on the oxidation

663 of organic carbon, and thus a positive carbon-nitrate relationship. We note that such
664 inferences should be interpreted as speculative, given that controlled experiments
665 were not conducted.
666
667 *Conclusions*
668 Our findings support the hypothesis that permafrost microbial communities
669 are shaped by factors that are distinct from those governing non-permafrost soil
670 communities. We found that microbial distributions in permafrost are driven
671 primarily by dispersal limitation imposed by frozen soil and deterministic selection
672 arising from thermodynamic constraints of the permafrost environment. This
673 contrasts sharply with non-permafrost soil communities, which are driven primarily
674 by soil pH [11-13]. These findings underscore the need for different mechanistic models
675 predicting microbial community characteristics in permafrost and non-permafrost
676 soils, given the different processes governing these systems. Our findings suggest
677 that predictive models of permafrost community composition will need to account
678 for organic carbon thermodynamics, organic acid concentrations, and redox
679 conditions, which may be informed by knowledge of landscape history. However,
680 efforts to accurately predict community composition at the landscape-scale based
681 solely on environmental characteristics may be limited due to the strong influence
682 of dispersal limitation.
683 Our findings additionally suggest that changes in permafrost microbial
684 community structure and function are likely to be drastic in response to thaw, as
685 hydrologic changes mobilize organisms and nutrients, thereby relieving the primary

686 constraints on communities. Community responses to change are also likely to be
687 highly varied across landscapes, given the environmental and microbiological
688 heterogeneity of permafrost-affected soils. Identifying how pre-thaw environmental
689 and community characteristics influence post-thaw responses will be essential for
690 accurately predicting ecosystem level responses to environmental change.
691

692 **References**

693 1 Van Everdingen, R. O. *Multi-language Glossary of Permafrost and Related*
694 *Ground-ice Terms in Chinese, English, French, German, Icelandic, Italian,*
695 *Norwegian, Polish, Romanian, Russian, Spanish, and Swedish*. (International
696 Permafrost Association, Terminology Working Group, 1998).

697 2 Burn, C. R. & Nelson, F. E. Comment on "A projection of severe near-surface
698 permafrost degradation during the 21st century" by David M. Lawrence and
699 Andrew G. Slater. *Geophysical Research Letters* **33**, L21503,
700 doi:10.1029/2006GL027077 (2006).

701 3 Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar
702 permafrost region. *Global Biogeochemical Cycles* **23**, GB2023,
703 doi:10.1029/2008GB003327 (2009).

704 4 Zhang, T., Barry, R. G., Knowles, K., Heginbottom, J. A. & Brown, J. Statistics
705 and characteristics of permafrost and ground-ice distribution in the Northern
706 Hemisphere. *Polar Geography* **31**, 47-68, doi:10.1080/10889370802175895
707 (2008).

708 5 Bockheim, J. G. Permafrost distribution in the southern circumpolar region
709 and its relation to the environment: A review and recommendations for
710 further research. *Permafrost and Periglacial Processes* **6**, 27-45,
711 doi:10.1002/ppp.3430060105 (1995).

712 6 Ping, C. L., Jastrow, J. D., Jorgenson, M. T., Michaelson, G. J. & Shur, Y. L.
713 Permafrost soils and carbon cycling. *Soil* **1**, 147-171, doi:10.5194/soil-1-147-
714 2015 (2015).

715 7 Jansson, J. K. & Tas, N. The microbial ecology of permafrost. *Nature Reviews*
716 *Microbiology* **12**, 414-425, doi:10.1038/nrmicro3262 (2014).
717 8 Schuur, E. A. G. *et al.* Vulnerability of Permafrost Carbon to Climate Change:
718 Implications for the Global Carbon Cycle. *BioScience* **58**, 701-714,
719 doi:10.1641/B580807 (2008).
720 9 Vaughan, D. G. *et al.* in *Climate Change 2013: The Physical Science Basis.*
721 *Contribution of Working Group I to the Fifth Assessment Report of the*
722 *Intergovernmental Panel on Climate Change* (eds T.F. Stocker *et al.*) Ch. 4,
723 317-382 (Cambridge University Press, 2013).
724 10 Mackelprang, R., Saleska, S. R., Jacobsen, C. S., Jansson, J. K. & Taş, N.
725 Permafrost Meta-Omics and Climate Change. *Annual Review of Earth and*
726 *Planetary Sciences* **44**, 439-462 (2016).
727 11 Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based
728 assessment of soil pH as a predictor of soil bacterial community structure at
729 the continental scale. *Applied and Environmental Microbiology* **75**, 5111-5120
730 (2009).
731 12 Chu, H. *et al.* Soil bacterial diversity in the Arctic is not fundamentally
732 different from that found in other biomes. *Environmental Microbiology* **12**,
733 2998-3006 (2010).
734 13 Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial
735 communities. *Proceedings of the National Academy of Sciences of the United*
736 *States of America* **103**, 626-631 (2006).

14 Cleveland, C. C., Nemergut, D. R., Schmidt, S. K. & Townsend, A. R. Increases in soil respiration following labile carbon additions linked to rapid shifts in soil microbial community composition. *Biogeochemistry* **82**, 229-240, doi:10.1007/s10533-006-9065-z (2007).

15 Ramirez, K. S., Craine, J. M. & Fierer, N. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Global Change Biology* **18**, 1918-1927 (2012).

16 Brockett, B. F., Prescott, C. E. & Grayston, S. J. Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. *Soil Biology and Biochemistry* **44**, 9-20 (2012).

17 Cruz-Martínez, K. *et al.* Effect of rainfall-induced soil geochemistry dynamics on grassland soil microbial communities. *Applied and Environmental Microbiology* **78**, 7587-7595 (2012).

18 Zhao, C. *et al.* Soil microbial community composition and respiration along an experimental precipitation gradient in a semiarid steppe. *Scientific Reports* **6**, 24317, doi:10.1038/srep24317 (2016).

19 Steven, B. *et al.* Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiology Ecology* **59**, 513-523, doi:10.1111/j.1574-6941.2006.00247.x (2007).

20 Tuorto, S. J. *et al.* Bacterial genome replication at subzero temperatures in permafrost. *The ISME Journal* **8**, 139 (2014).

760 21 Trivedi, P., Anderson, I. C. & Singh, B. K. Microbial modulators of soil carbon
761 storage: integrating genomic and metabolic knowledge for global prediction.
762 *Trends in Microbiology* **21**, 641-651 (2013).
763 22 Wieder, W. R. *et al.* Explicitly representing soil microbial processes in Earth
764 system models. *Global Biogeochemical Cycles* **29**, 1782-1800 (2015).
765 23 Coolen, M. J., van de Giessen, J., Zhu, E. Y. & Wuchter, C. Bioavailability of soil
766 organic matter and microbial community dynamics upon permafrost thaw.
767 *Environmental Microbiology* **13**, 2299-2314 (2011).
768 24 Ernakovich, J. G., Wallenstein, M. D. & Calderón, F. J. Chemical Indicators of
769 Cryoturbation and Microbial Processing throughout an Alaskan Permafrost
770 Soil Depth Profile. *Soil Science Society of America Journal* **79**, 783-793,
771 doi:10.2136/sssaj2014.10.0420 (2015).
772 25 Lee, H., Schuur, E. A., Inglett, K. S., Lavoie, M. & Chanton, J. P. The rate of
773 permafrost carbon release under aerobic and anaerobic conditions and its
774 potential effects on climate. *Global Change Biology* **18**, 515-527 (2012).
775 26 Waldrop, M. P. *et al.* Molecular investigations into a globally important
776 carbon pool: permafrost-protected carbon in Alaskan soils. *Global Change*
777 *Biology* **16**, 2543-2554, doi:10.1111/j.1365-2486.2009.02141.x (2010).
778 27 Coolen, M. J. L. & Orsi, W. D. The transcriptional response of microbial
779 communities in thawing Alaskan permafrost soils. *Frontiers in Microbiology*
780 **6**, doi:10.3389/fmicb.2015.00197 (2015).

781 28 Mackelprang, R. *et al.* Metagenomic analysis of a permafrost microbial
782 community reveals a rapid response to thaw. *Nature* **480**, 368-371,
783 doi:10.1038/nature10576 (2011).
784 29 McCalley, C. *et al.* Methane dynamics regulated by microbial community
785 response to permafrost thaw. *Nature* **514**, 478-481 (2014).
786 30 Mondav, R. *et al.* Discovery of a novel methanogen prevalent in thawing
787 permafrost. *Nature Communications* **5**, 3212, doi:10.1038/ncomms4212
788 (2014).
789 31 Hultman, J. *et al.* Multi-omics of permafrost, active layer and thermokarst bog
790 soil microbiomes. *Nature* **521**, 208-212, doi:10.1038/nature14238 (2015).
791 32 Burrows, S. N. *et al.* Application of Geostatistics to Characterize Leaf Area
792 Index (LAI) from Flux Tower to Landscape Scales Using a Cyclic Sampling
793 Design. *Ecosystems* **5**, 0667-0679, doi:10.1007/s10021-002-0110-z (2002).
794 33 Stookey, L. L. Ferrozine---a new spectrophotometric reagent for iron.
795 *Analytical Chemistry* **42**, 779-781 (1970).
796 34 Tfaily, M. M. *et al.* Advanced solvent based methods for molecular
797 characterization of soil organic matter by high-resolution mass
798 spectrometry. *Analytical Chemistry* **87**, 5206-5215 (2015).
799 35 Tfaily, M. M. *et al.* Sequential extraction protocol for organic matter from soils
800 and sediments using high resolution mass spectrometry. *Analytica Chimica*
801 *Acta* **972**, 54-61 (2017).

802 36 Kujawinski, E. B. & Behn, M. D. Automated analysis of electrospray ionization
803 Fourier transform ion cyclotron resonance mass spectra of natural organic
804 matter. *Analytical Chemistry* **78**, 4363-4373 (2006).
805 37 Minor, E. C., Steinbring, C. J., Longnecker, K. & Kujawinski, E. B.
806 Characterization of dissolved organic matter in Lake Superior and its
807 watershed using ultrahigh resolution mass spectrometry. *Organic*
808 *Geochemistry* **43**, 1-11 (2012).
809 38 LaRowe, D. E. & Van Cappellen, P. Degradation of natural organic matter: a
810 thermodynamic analysis. *Geochimica et Cosmochimica Acta* **75**, 2030-2042
811 (2011).
812 39 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on
813 the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**, 1621-1624
814 (2012).
815 40 Aronesty, E. Comparison of sequencing utility programs. *The Open*
816 *Bioinformatics Journal* **7**, 1-8 (2013).
817 41 BBMap: a fast, accurate, splice-aware aligner (2014).
818 42 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
819 *Bioinformatics* **26**, 2460-2461 (2010).
820 43 RStudio: Integrated Development for R (RStudio, Inc., Boston, MA, 2015).
821 44 Vegan: community ecology package v. R package version 2.4-3. (2017).
822 45 Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W. & Heuvelink, G. B. M. Real-
823 time automatic interpolation of ambient gamma dose rates from the Dutch

radioactivity monitoring network. *Computers & Geosciences* **35**, 1711-1721, doi:doi:10.1016/j.cageo.2008.10.011 (2009).

46 Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**, 51-68 (2002).

47 Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**, 1826-1832, doi:10.1890/03-3111 (2004).

48 Borcard, D., Gillet, F. & Legendre, P. *Numerical Ecology with R*. (Springer-Verlag, 2011).

49 Dini-Andreote, F., Stegen, J. C., van Elsas, J. D. & Salles, J. F. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1326-E1332, doi:10.1073/pnas.1414261112 (2015).

50 Stegen, J. C., Lin, X., Fredrickson, J. K. & Konopka, A. E. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology* **6**, doi:10.3389/fmicb.2015.00370 (2015).

51 leaps: Regression Subset Selection (2017).

52 Structural Equation Models (2017).

53 Shur, Y., Hinkel, K. M. & Nelson, F. E. The transient layer: implications for geocryology and climate-change science. *Permafrost and Periglacial Processes* **16**, 5-17, doi:10.1002/ppp.518 (2005).

847 54 McMahon, P. B. & Chapelle, F. H. Microbial production of organic acids in
848 aquitard sediments and its role in aquifer geochemistry. *Nature* **349**, 233-
849 235 (1991).
850 55 Duddleston, K. N., Kinney, M. A., Kiene, R. P. & Hines, M. E. Anaerobic
851 microbial biogeochemistry in a northern bog: Acetate as a dominant
852 metabolic end product. *Global Biogeochemical Cycles* **16**, 11-11-11-19,
853 doi:10.1029/2001GB001402 (2002).
854 56 Hines, M. E., Duddleston, K. N. & Kiene, R. P. Carbon flow to acetate and C1
855 compounds in northern wetlands. *Geophysical Research Letters* **28**, 4251-
856 4254, doi:10.1029/2001GL012901 (2001).
857 57 Tas, N. *et al.* Impact of fire on active layer and permafrost microbial
858 communities and metagenomes in an upland Alaskan boreal forest. *The ISME*
859 *Journal* **8**, 1904-1919, doi:10.1038/ismej.2014.36 (2014).
860 58 Bergmann, G. T. *et al.* The under-recognized dominance of Verrucomicrobia
861 in soil bacterial communities. *Soil Biology & Biochemistry* **43**, 1450-1455,
862 doi:10.1016/j.soilbio.2011.03.012 (2011).
863 59 Vellend, M. Conceptual Synthesis in Community Ecology. *The Quarterly*
864 *Review of Biology* **85**, 183-206, doi:10.1086/652373 (2010).
865 60 Stegen, J. C. *et al.* Quantifying community assembly processes and identifying
866 features that impose them. *The ISME Journal* **7**, 2069-2079,
867 doi:10.1038/ismej.2013.93 (2013).
868 61 Lipson, D. A., Jha, M., Raab, T. K. & Oechel, W. C. Reduction of iron (III) and
869 humic substances plays a major role in anaerobic respiration in an Arctic

peat soil. *Journal of Geophysical Research: Biogeosciences* **115**, G00I06, doi:10.1029/2009JG001147 (2010).

62 Lipson, D. A., Raab, T. K., Goria, D. & Zlamal, J. The contribution of Fe(III) and humic acid reduction to ecosystem respiration in drained thaw lake basins of the Arctic Coastal Plain. *Global Biogeochemical Cycles* **27**, 399-409, doi:10.1002/gbc.20038 (2013).

63 Miller, K. E., Lai, C.-T., Friedman, E. S., Angenent, L. T. & Lipson, D. A. Methane suppression by iron and humic acids in soils of the Arctic Coastal Plain. *Soil Biology and Biochemistry* **83**, 176-183, doi:doi:10.1016/j.soilbio.2015.01.022 (2015).

64 Barberán, A. *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 5756-5761, doi:10.1073/pnas.1420815112 (2015).

65 Treves, D. S., Xia, B., Zhou, J. & Tiedje, J. M. A Two-Species Test of the Hypothesis That Spatial Isolation Influences Microbial Diversity in Soil. *Microbial Ecology* **45**, 20-28, doi:10.1007/s00248-002-1044-x (2003).

66 Austin, A. T. *et al.* Water pulses and biogeochemical cycles in arid and semiarid ecosystems. *Oecologia* **141**, 221-235, doi:10.1007/s00442-004-1519-1 (2004).

67 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* **66**, 265-286, doi:10.1023/B:BIOG.0000005341.19412.7b (2003).

893 68 Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification
894 of soil bacteria. *Ecology* **88**, 1354-1364, doi:10.1890/05-1839 (2007).
895 69 Hartmann, A., Schmid, M., Tuinen, D. v. & Berg, G. Plant-driven selection of
896 microbes. *Plant and Soil* **321**, 235-257, doi:10.1007/s11104-008-9814-y
897 (2009).
898
899

Acknowledgments

This research was supported by the Microbiomes in Transition Initiative at Pacific Northwest National Lab (PNNL), a multiprogram 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. 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.

Author Contributions

This work was conceived by EMB, JKJ, and JCS. EMB, DWK, EBR, and SJF completed sample processing and laboratory analyses. Processing of amplicon sequencing data was completed by JMB. MMT and RKC performed FTICR-MS analyses and data processing. Statistical analyses were completed by EMB, JCS, and LMB. EMB wrote the manuscript, with input from all co-authors.

Competing Financial Interests

The authors declare no competing financial interests.