

# Northumbria Research Link

Citation: Malard, Lucie, Anwar, Muhammad Zohaib, Jacobsen, Carsten S. and Pearce, David (2021) Influence of Spatial Scale on Structure of Soil Bacterial Communities across an Arctic Landscape. *Applied and Environmental Microbiology*, 87 (5). e02220-20. ISSN 0099-2240

Published by: American Society for Microbiology

URL: <http://doi.org/10.1128/AEM.02220-20> <<http://doi.org/10.1128/AEM.02220-20>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/45116/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



# Influence of Spatial Scale on Structure of Soil Bacterial Communities across an Arctic Landscape

 Lucie A. Malard,<sup>a\*</sup> Muhammad Z. Anwar,<sup>b\*</sup> Carsten S. Jacobsen,<sup>b</sup> David A. Pearce<sup>a</sup>

<sup>a</sup>Faculty of Health and Life Sciences, Northumbria University, Newcastle-upon-Tyne, United Kingdom

<sup>b</sup>Department of Environmental Sciences, Aarhus University, Roskilde, Denmark

**ABSTRACT** Bacterial community composition is largely influenced by environmental factors, and this applies to the Arctic region. However, little is known about the role of spatial factors in structuring such communities. In this study, we evaluated the influence of spatial scale on bacterial community structure across an Arctic landscape. Our results showed that spatial factors accounted for approximately 10% of the variation at the landscape scale, equivalent to observations across the whole Arctic region, suggesting that while the role and magnitude of other processes involved in community structure may vary, the role of dispersal may be stable globally in the region. We assessed dispersal limitation by identifying the spatial autocorrelation distance which would be required in order to obtain fully independent samples as approximately 60 m, and this may inform future sampling strategies in the region. Finally, indicator taxa with strong statistical correlations with environmental variables were identified. However, we showed that these strong taxon-environment associations may not always be reflected in the geographical distribution of these taxa.

**IMPORTANCE** The significance of this study is 3-fold. It investigated the influence of spatial scale on the soil bacterial community composition across a typical Arctic landscape and demonstrated that conclusions reached when examining the influence of specific environmental variables on bacterial community composition are dependent upon the spatial scales over which they are investigated. This study identified a dispersal limitation (spatial autocorrelation) distance of approximately 60 m, required to obtain samples with fully independent bacterial communities and, therefore, should serve to inform future sampling strategies in the region and potentially elsewhere. The work also showed that strong taxon-environment statistical associations may not be reflected in the observed landscape distribution of the indicator taxa.

**KEYWORDS** Arctic soil, autocorrelation distance, bacterial diversity, biogeography, environmental factors, spatial scale

Significant spatial structuring of soil microorganisms has been demonstrated at micro (micrometers to millimeters) (1), plot (centimeters to a few meters) (1), landscape (hundreds of meters) (2), regional (kilometers) (3), national (4, 5), continental (6), and global (7–9) scales. Hence, the scale of investigation is a key parameter to take into account in studies of bacterial biogeography. Martiny et al. (10) further demonstrated the importance of spatial scale for environmental factors identified as influencing community composition in temperate soils. They found key environmental drivers differed across spatial scales—ammonia-oxidizing bacterial (AOB) community composition was dependent on distance, moisture, and vegetation cover at the plot scale; however, at the regional scale, diversity was mainly influenced by water temperature, air temperature, and moisture, while nitrate concentration and air temperature were predominant at the continental scale. Finally, when all scales were considered together, overall key drivers were geographic distance, sediment moisture, air temperature, and

**Citation** Malard LA, Anwar MZ, Jacobsen CS, Pearce DA. 2021. Influence of spatial scale on structure of soil bacterial communities across an Arctic landscape. *Appl Environ Microbiol* 87:e02220-20. <https://doi.org/10.1128/AEM.02220-20>.

**Editor** Haruyuki Atomi, Kyoto University

**Copyright** © 2021 Malard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Lucie A. Malard, [lucie.malard@unil.ch](mailto:lucie.malard@unil.ch), or David A. Pearce, [david.pearce@northumbria.ac.uk](mailto:david.pearce@northumbria.ac.uk).

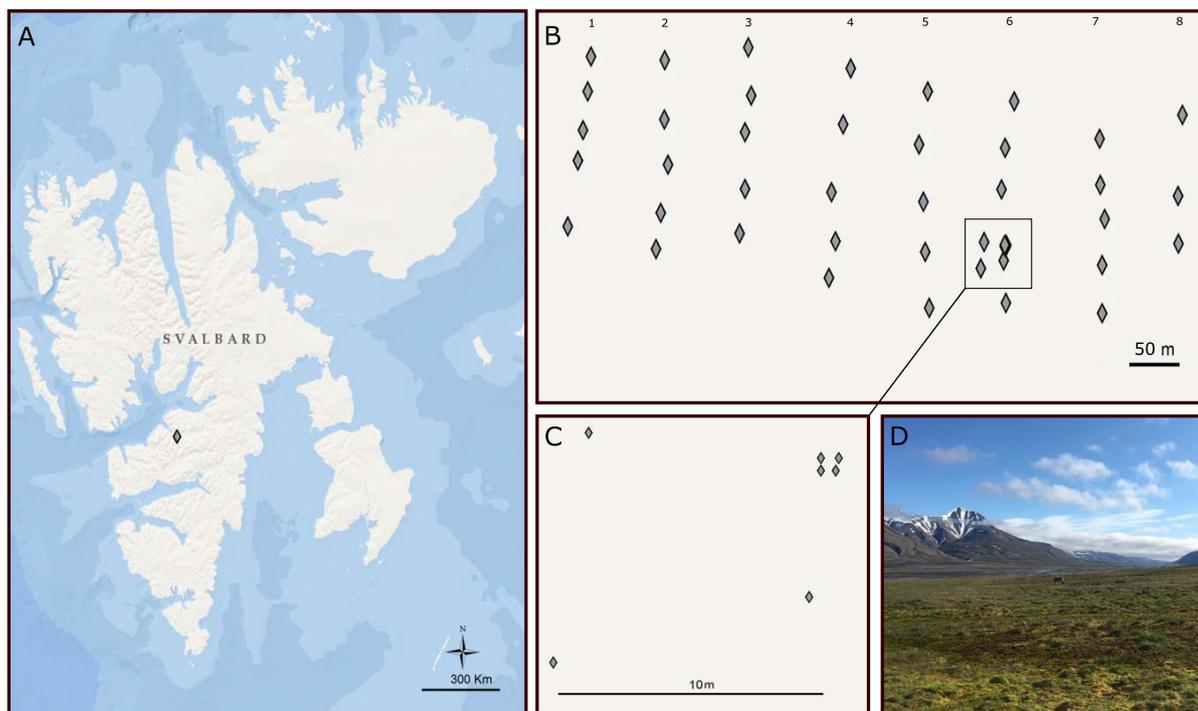
\* Present address: Lucie A. Malard, Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland; Muhammad Z. Anwar, Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

**Received** 9 September 2020

**Accepted** 10 December 2020

**Accepted manuscript posted online** 23 December 2020

**Published** 12 February 2021



**FIG 1** (A) Map of sampling sites in Svalbard, created using Arcmap version 10.7. (B) Sampling design in 8 transects in Adventdalen. (C) Smaller scale samples on transect 6. (D) View of Adventdalen.

vegetation cover. However, most biogeographical studies investigate communities at only one spatial scale (see references 4, 7, and 9 for further examples). The landscape scale (a few hundred meters to a few kilometers) is considered highly relevant for studies of bacterial distribution patterns, as it is the scale of human activities (at which agricultural practices and land management are integrated). Hence, the majority of studies at that scale investigate human-impacted landscapes (see references 2, 3, 11, 12, and 13 for further examples), with only a few studies describing Arctic communities from a few meters to 3 km (14–16).

The first aim of this study was to evaluate the influence of the spatial scale on bacterial community structure (see Fig. S1 in the supplemental material) across an Arctic landscape (Fig. 1). Indeed, while the role of environmental parameters such as pH (17, 18), total organic content (TOC) (19), moisture content (20), and C/N ratio (21) in community composition in the Arctic has been demonstrated, much less is known about the influence of spatial parameters (19). However, determining the influence of environmental factors on communities remains an essential step to avoid overestimating the role of the spatial scale. In addition to providing a better understanding of the environmental factors influencing community structure, investigating multiple scales provides better knowledge of the spatial structure, which facilitates the development of sampling strategies where samples are collected beyond the spatial autocorrelation distance and are, therefore, truly independent (22). As autocorrelation distances have been identified from micrometers to kilometers (22–25), with the potential of nested scales of variability (26) and site-to-site variation, no standardized protocol exists for soil sampling for metabarcoding studies (27, 28). Therefore, the second aim was to determine the minimum distance required to obtain independent soil samples in the region (Fig. S1), which may inform future sampling strategies in the Arctic. Finally, the last aim was to identify indicator taxa which were closely associated with environmental variables and map their spatial distribution across the landscape (Fig. S1). Previous studies have attempted to identify indicator taxa that could be used for environmental monitoring (for example, see references 29 and 30 [rivers] and 31 [soils]). As indicator

**TABLE 1** Relative influence of environmental factors on bacterial community structure<sup>a</sup>

Variable	R <sup>2</sup>	Pr(>F)
TOC	0.089	0.001***
pH	0.070	0.001***
Cond	0.059	0.001***
Al	0.041	0.001***
As	0.041	0.001***
Br	0.024	0.001***
La	0.022	0.001***
Y	0.021	0.002**
Ca	0.018	0.003**
Cl	0.018	0.001***
Cs	0.018	0.001***
Pb	0.018	0.001***
Sr	0.018	0.002**
S	0.016	0.001***
Cu	0.015	0.001***
Te	0.015	0.002**
Ba	0.014	0.003**
In	0.014	0.002**
Nb	0.014	0.004**
Nd	0.014	0.008**
Si	0.014	0.004**
Fe	0.013	0.002**
I	0.013	0.006**
Mn	0.013	0.009**
Th	0.013	0.005**
Ag	0.012	0.007**
Mo	0.012	0.013*
Sb	0.012	0.010**
Cd	0.011	0.023*
Ta	0.011	0.016*
Tl	0.011	0.021*
Zr	0.011	0.012*
Zn	0.010	0.031*
Ge	0.009	0.046*
Sn	0.009	0.036*
Residuals	0.269	NA

<sup>a</sup>Calculated by PERMANOVA using the adonis function. Pr(>F), P value of ANOVA (output via R); NA, not applicable.

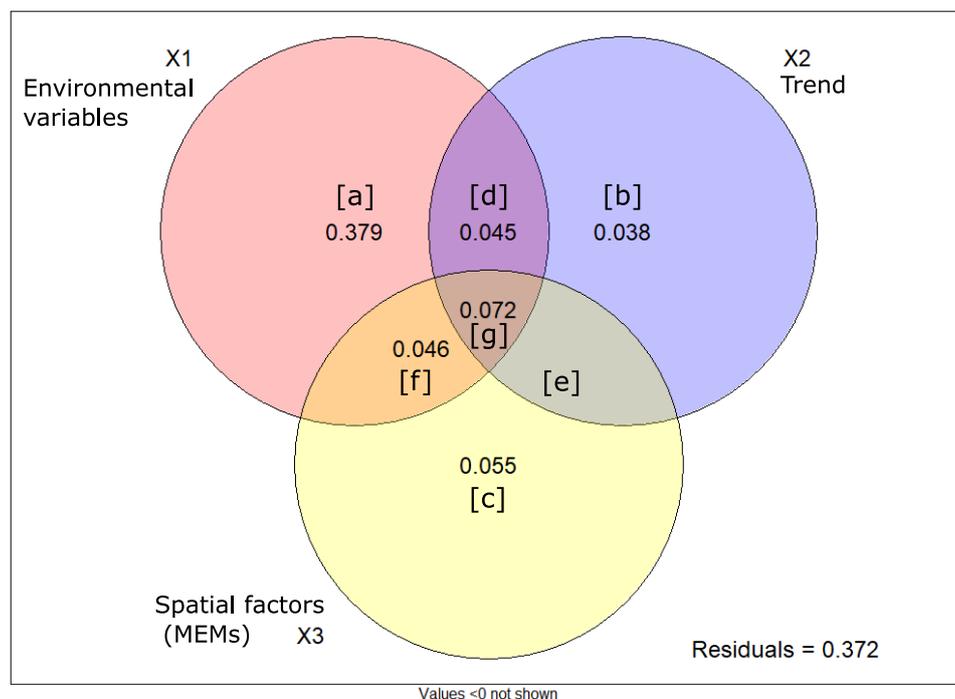
\*, 0.05 > P > 0.01; \*\*, 0.01 > P > 0.001; \*\*\*, P < 0.001.

taxa (32) highlight operational taxonomic units (OTUs) with strong environmental associations that may respond to ecological change, we expected their distribution to closely follow that of environmental parameters.

## RESULTS

**Environmental factors.** Results showed that all 35 environmental variables had a significant impact on bacterial community structure, with approximately 73% of the variance explained by environmental factors (Table 1). Overall, five key factors (TOC, pH, conductivity, aluminum levels, and arsenic levels) had the most influence on bacterial community dissimilarity, explaining 30% of variation in total. While all other environmental factors individually explained between 0.9% and 2.4% of the variation, the combined soil elemental composition (excluding pH, conductivity, and TOC) accounted for 51.5% of the total variation in bacterial community composition.

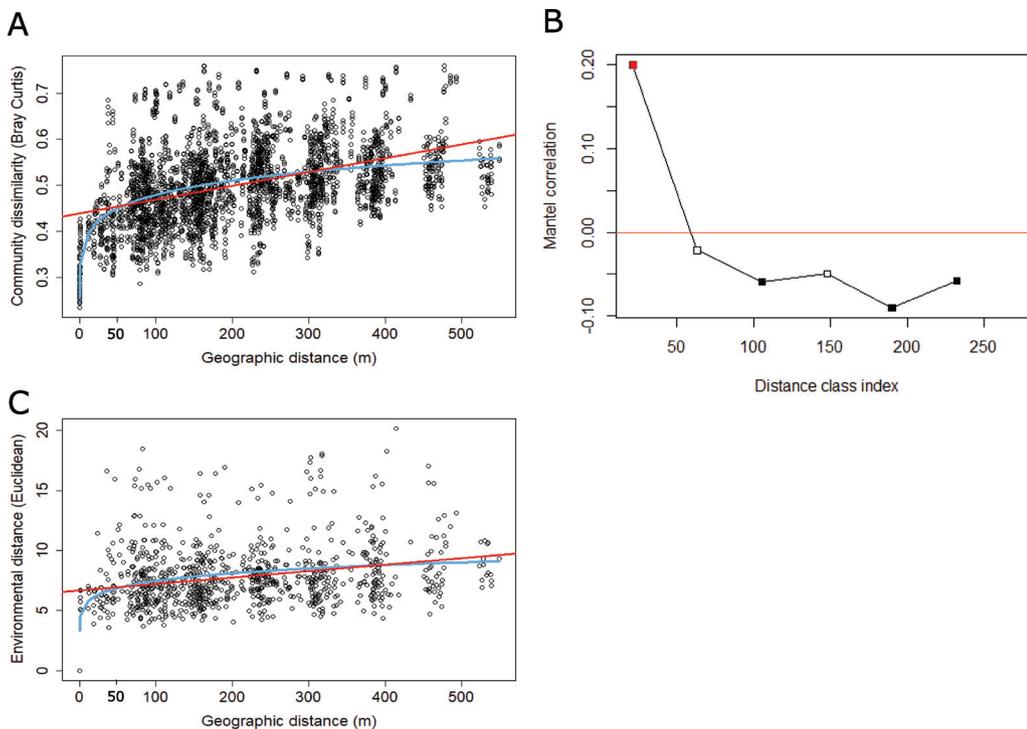
**Variation partitioning.** A total of 9 distance-based Moran's eigenvector map (dbMEM) vectors were built using (x,y) geographic coordinates, and after forward selection, five dbMEMs were identified as significantly impacting bacterial community diversity and used in subsequent analyses. The variation partitioning analysis differentiated the effect of environmental factors, linear trend, and spatial vectors on community composition (Fig. 2). The environmental fraction X1 explained 73% of the variance



**FIG 2** Venn diagram illustrating the results of the variation partitioning analysis on the influence of environmental variables and spatial factors on bacterial community composition. Results of each partition can be multiplied by 100 for the percentage of variation explained and are detailed in Table S2.

(Table S1), equal to the finding by the adonis function and confirming the success of the variation partitioning analysis. Using the adjusted  $R^2$  values only as they accounted for the number of variables in the model, environmental factors explained 54% of the variance, of which 38% were not spatially structured (fraction a). The spatial component (X2 + X3) explained 25.6% of the variation, of which 16.3% could be explained by induced spatial dependence. This was illustrated by fractions d, f, and g, which represented spatially structured environmental variables where the spatial structure of these environmental variables induced a similar spatial structure in the response data, highlighting the need to evaluate the influence of the environment on communities. The remaining 9.3% of the spatial component represented spatial autocorrelation. The linear trend accounted for 3.8% of the variance (fraction b), while spatial vectors explained 5.5% of the variation. Fraction e had a negative  $R^2$  and could be considered null, as prescribed in reference 33. Each fraction (X1, X2, and X3) was tested individually and was significant (analysis of variance [ANOVA];  $P < 0.001$ ). In total, 62.8% of the bacterial community dissimilarity could be explained by environmental and spatial factors, while the remaining 37.2% of the variance could not be explained by the variables measured in this study.

**Spatial scale and autocorrelation.** The distance-decay curve illustrated the increase in community dissimilarity with increasing distance (Fig. 3A). The power model was better fitted ( $R^2 = 0.2261$ ;  $P = 0.005$ ) than the linear regression ( $R^2 = 0.1844$ ;  $P < 2.2 \times 10^{-16}$ ). Spatial autocorrelation was visualized on the distance-decay curve (Fig. 3), where geographically close communities were more similar up to 60 m. This was illustrated with the power model on the distance-decay curve, where the blue curve begins to plateau (Fig. 3A). To further characterize the spatial autocorrelation distance, a Mantel correlogram was used (Fig. 3B) to compute the Mantel statistic between the geographic distance and bacterial community dissimilarity distance (Bray-Curtis). The spatial autocorrelation was positive for the first distance class of 21 m, indicating that the bacterial communities were more similar than expected by chance. The second distance class of 63 m displayed no spatial autocorrelation, indicating random



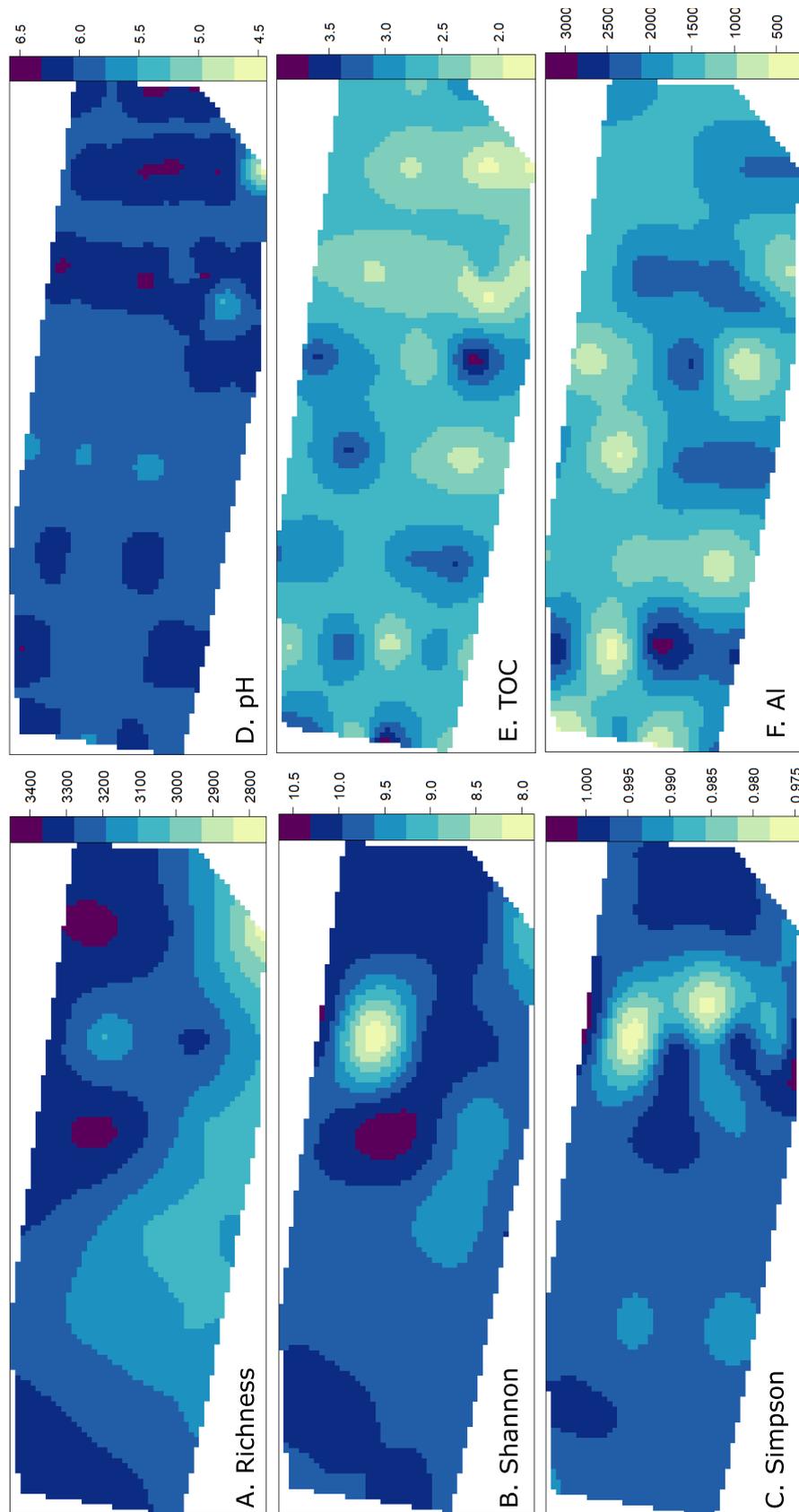
**FIG 3** (A) Distance-decay curve illustrating the increase in bacterial community dissimilarity with increasing geographic distance. (B) Mantel correlogram of spatial autocorrelation illustrating the dispersal limitation. The red square indicates positive significant autocorrelation, which was identified only in the first distance class (0 to 21 m). Beyond 60 m, the autocorrelation was either negative (black squares) or not significant (white squares). (C) Distance-decay curve illustrating increasing environmental variation with increasing geographic distance. The red curve illustrates the linear regression, and the blue curve is the power model.

distribution beyond 63 m. Other distance classes presented negative autocorrelations, indicating that these bacterial communities were more different than expected by chance.

Geography also had some influence on environmental conditions, with sites closer together being more similar. The spatial autocorrelation of environmental variables was first visualized in Fig. 3C, where geographically close sites were geochemically similar within 25 m. However, beyond approximately 25 m, sites equally close or far could present similar environmental conditions, as illustrated by the autocorrelation distance (Fig. 3C). This was also illustrated by the weak linear regression ( $R^2 = 0.019$ ;  $P < 2.2 \times 10^{-16}$ ) and the best-fitted power model ( $R^2 = 0.087$ ;  $P = 0.005$ ). Spatial autocorrelation was further tested for each individual variable using the semivariograms produced prior to kriging. As semivariograms are specific to each variable, the spatial autocorrelation distances were unique to each parameter. All the semivariograms produced prior to Kriging indicated positive autocorrelations oscillating between 1 m and 100 m, depending on the variable tested, further illustrating the importance of the scale of investigation (Fig. S2).

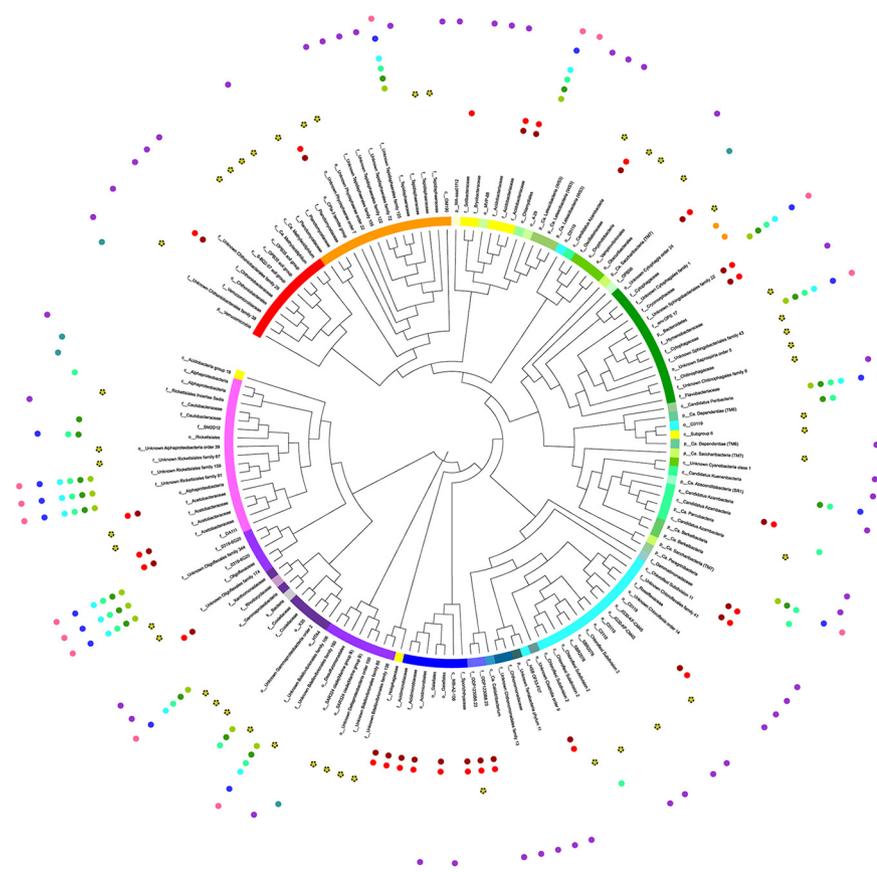
**Spatial distribution across the landscape.** Using an ordinary kriging method and after examining the semivariograms, the spatial distribution of alpha diversity and key environmental variables was mapped across the landscape (Fig. 4). The bacterial richness, diversity, and evenness changed across the landscape (Fig. 4A to C), and kriged maps illustrated the relationships between diversity, evenness, and richness. Overall, low richness indicated low diversity and low evenness, further observed using linear models (Fig. S3). The kriged maps of alpha diversity and environmental variables showed the strong heterogeneity at the landscape scale, with changes from high to low concentrations within just a few meters (Fig. 4D to F).

**Indicator taxa.** The indicator species analysis identified 163 true specialist (statistic  $> 0.98$ ) OTUs associated with 12 environmental variables. Indicator taxa were generally



**FIG 4** Kriged maps of the spatial distribution across the landscape showing the heterogeneity of (A) richness, (B) Shannon index, (C) Simpson index, (D) pH, (E) total organic carbon, and (F) aluminum. The (Continued on next page)

Downloaded from <http://aem.asm.org/> on May 26, 2021 by guest



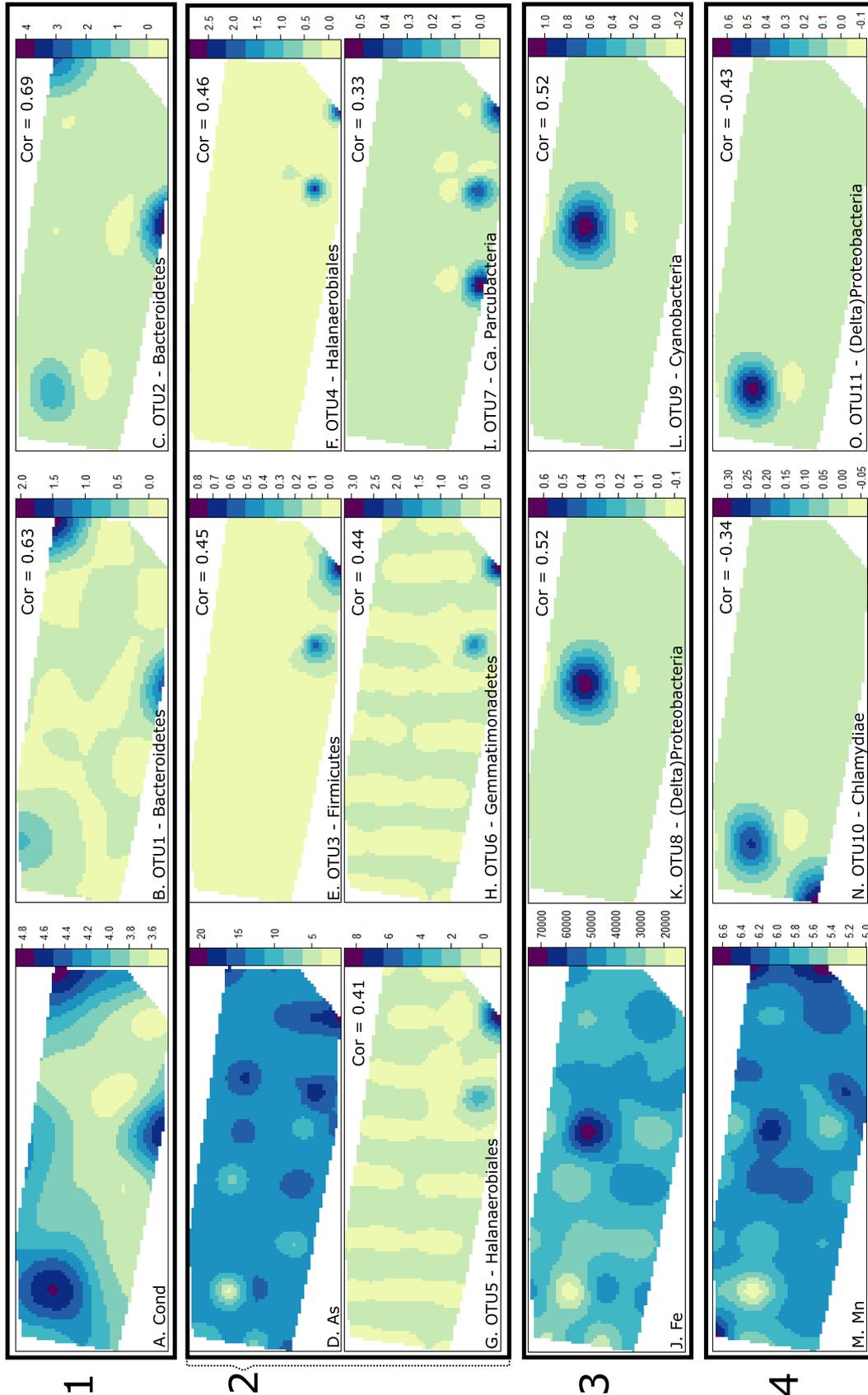
Classification			Element	
<span style="color: red;">■</span> Verrucomicrobia	<span style="color: green;">■</span> Bacteroidetes	<span style="color: blue;">■</span> Armatimonadetes	<span style="color: darkred;">●</span> As	<span style="color: lightgreen;">●</span> Nb
<span style="color: orange;">■</span> Planctomycetes	<span style="color: lightgreen;">■</span> Peregrinibacteria	<span style="color: lightblue;">■</span> Atribacteria	<span style="color: red;">●</span> Ba	<span style="color: cyan;">●</span> Pb
<span style="color: yellow;">■</span> WA	<span style="color: green;">■</span> TM6	<span style="color: purple;">■</span> Halanaerobiales	<span style="color: orange;">●</span> Cond	<span style="color: teal;">●</span> Sr
<span style="color: yellowgreen;">■</span> Acidobacteria	<span style="color: lightgreen;">■</span> SR1	<span style="color: blue;">■</span> Actinobacteria	<span style="color: black;">●</span> Fe	<span style="color: blue;">●</span> Y
<span style="color: lightgreen;">■</span> Elusimicrobia	<span style="color: green;">■</span> Parcubacteria	<span style="color: purple;">■</span> Deltaproteobacteria	<span style="color: yellowgreen;">●</span> Mn	<span style="color: purple;">●</span> Zn
<span style="color: lightgreen;">■</span> Chlamydiae	<span style="color: green;">■</span> Berkelbacteria	<span style="color: darkpurple;">■</span> Gammaproteobacteria	<span style="color: green;">●</span> Mo	<span style="color: pink;">●</span> Zr
<span style="color: lightgreen;">■</span> WS3	<span style="color: lightblue;">■</span> Gemmatimonadetes	<span style="color: grey;">■</span> Unclassified		
<span style="color: green;">■</span> Cyanobacteria	<span style="color: cyan;">■</span> Chloroflexi	<span style="color: lightpurple;">■</span> Betaproteobacteria		
<span style="color: yellowgreen;">■</span> TM7	<span style="color: grey;">■</span> Firmicutes	<span style="color: pink;">■</span> Alphaproteobacteria		
<span style="color: lightgreen;">■</span> Chlorobi	<span style="color: darkgrey;">■</span> Terrabacteria			

**FIG 5** Phylogenetic tree of indicator taxa associated with environmental variables showing the high phylogenetic diversity. Colored bands illustrate the taxonomy of each OTU at the phylum level (or class level for *Proteobacteria*); labels indicate the taxonomy down to the family level if available. Colored points indicate the associated element.

associated with the highest concentration of each element. The phylogenetic tree specific to indicator taxa illustrated the high taxonomic diversity of indicator taxa (Fig. 5); however, Fig. 6 demonstrates that identified indicator taxa do not necessarily follow environmental gradients as they are expected to. Of the four key factors (excluding pH) influencing bacterial communities (Table 1), only conductivity and arsenic had some associated indicator taxa. Indicators of conductivity (Cond) were restricted to

**FIG 4** Legend (Continued)

color bars in panels A, B, and C indicate values of alpha diversity, while those for the environmental variables indicate element concentrations (units for each variable are given in Table S2, taking into account data transformations).



**FIG 6** Spatial distribution across the landscape using a kriged map and illustrating the heterogeneous distribution. The color bars for environmental variables indicate element concentrations (Table S2 gives units, considering data (Continued on next page)

two OTUs associated with high conductivity, both members of the *Bacteroidetes* classified in the order *Cytophagales* (Fig. 5). Peaks of high conductivity were visualized in Fig. 6A and correlated with peaks in abundance of the two OTUs identified (Fig. 6B and C). Indicators of arsenic (As) were closely associated with barium (Ba) and were taxonomically diverse, with the majority classified as *Actinobacteria*, *Alphaproteobacteria*, *Chloroflexi*, *Halanaerobiales incertae sedis*, and *Firmicutes* (Fig. 5). Arsenic concentration appeared more homogeneous across the landscape (Fig. 6D), with an average concentration of 13 parts per million (ppm), a minimum of 1.81 ppm, and a maximum of 20.51 ppm. These indicator taxa of arsenic were all associated with high concentrations (Fig. 6E to I) and were also associated with high concentrations of barium in the soil.

Iron (Fe) and manganese (Mn) are both essential elements of soils. Iron concentration was highly heterogeneous across the landscape, with a strong peak in concentration at one site (Fig. 6J). This peak was reflected by the presence of unique indicator taxa whose abundance was closely related to this high concentration (Fig. 6K and L). Indicators of iron were diverse, with a large number of *Proteobacteria* (*Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*), *Chloroflexi*, *Bacteroidetes*, *Cyanobacteria*, *Planctomycetes*, and *Verrucomicrobia* (Fig. 5). On the other hand, manganese concentration was heterogeneous across the landscape (Fig. 6M), but unlike other indicator taxa, indicators of manganese were associated with low concentrations in the soil (Fig. 6N and O). The indicator taxa of manganese were predominantly classified as *Proteobacteria* (Fig. 5) and were also closely related to low concentrations of niobium (Nb), lead (Pb), and zirconium (Zr); however, they were associated with high concentrations of molybdenum (Mo). Indicator taxa of strontium (Sr) were limited to five OTUs, an unknown *Verrucomicrobia*, a "*Candidatus Saccharibacteria*" (TM7), a *Deltaproteobacteria*, and two *Alphaproteobacteria*, while indicators of zinc (Zn) were classified in all almost all phyla (Fig. 5), illustrating the wide array of specialist taxa associated with high concentrations of zinc.

## DISCUSSION

**Key environmental factors influencing bacterial communities.** Total organic carbon, pH, and conductivity were identified as the key drivers of bacterial diversity across the Arctic landscape and are also commonly identified in studies across the globe (8, 34–38). While pH was previously identified as the primary driver of bacterial diversity in Arctic soils across the whole Arctic region (19), here, at the landscape scale, TOC was identified as the primary factor influencing bacterial community structure and was tightly linked with soil moisture. Generally, soil organic carbon content increases with increasing precipitation and decreasing temperature (39). In the Arctic tundra, not only precipitation but also snowmelt and permafrost thaw have major impacts on soil moisture and hydrology across the landscape (40, 41). In this study, where pH was on average acidoneutral at  $6.05 \pm 0.36$  with very few acidic patches, but organic carbon content was very patchy (6% to 46%), the role of TOC in bacterial community structure is perhaps not surprising. However, it highlights the importance of investigating different spatial scales, as drivers at the global scale may not necessarily be the same across the landscape of interest.

Aluminum and arsenic were the fourth and fifth environmental variables accounting for the most variation in bacterial community structure (Table 1). Aluminum is one of the most abundant metals in the Earth's crust, and microorganisms continuously

### FIG 6 Legend (Continued)

transformations), while those for OTUs represent relative abundance. (Box 1) (A) Conductivity. (B) Phylum, *Bacteroidetes*; order, *Cytophagales*. (C) Phylum, *Bacteroidetes*; order, *Cytophagales*. (Box 2) (D) Arsenic. (E) Phylum, *Firmicutes*; order, unknown *Costridiales*. (F) Phylum, *Halanaerobiales incertae sedis*; order, *Halanaerobiales*. (G) Phylum, *Halanaerobiales incertae sedis*; order, *Halanaerobiales*. (H) Phylum, *Gemmatimonadetes*; order, *Gemmatimonadales*. (I) Phylum, "*Candidatus Parcubeacteria*"; class, "*Candidatus Azambacteria*." (Box 3) (J) Iron. (K) Phylum, *Proteobacteria* (*Deltaproteobacteria*); order, *Bdellovibrionales*. (L) Phylum, *Cyanobacteria*; order, *Oscillatoriales*. (Box 4) (M) Manganese. (N) Phylum, *Chlamydiae*; order, *Chlamydiales*. (O) Phylum, *Proteobacteria* (*Deltaproteobacteria*); order, *Oligoflexales*.

interact with aluminum in soils (42, 43). While aluminum lacks apparent biological function (42), the aluminum ion ( $\text{Al}^{3+}$ ) can be toxic to living organisms and is a function of the soil pH; the concentration of toxic  $\text{Al}^{3+}$  gradually increases as pH decreases from 6.2 (42, 43). Here, small pH changes but large aluminum concentration variations were observed across the landscape, which were not correlated with each other (linear regression:  $R^2 = 0.00069$ ;  $P = 0.81$ ). The toxicity of  $\text{Al}^{3+}$  may influence the bacterial community structure; however, the concentration of  $\text{Al}^{3+}$  ions was not measured.

Arsenic is ubiquitous in low abundance in the natural environment and recognized as one of the most toxic elements (44, 45). Here, a decrease in diversity and richness was observed with increasing arsenic concentrations, which likely reflects the toxic effect of oxyanions of arsenate on many bacteria, although some can use it as a terminal electron acceptor (44). As with  $\text{Al}^{3+}$ , the chemical concentration of the various forms of arsenic was not measured, and therefore, we cannot conclude that the toxicity has an influence on bacterial community structure, although it is a possibility. Indicator taxa associated with high concentrations of arsenic were diverse but dominated by *Actinobacteria* and *Proteobacteria*, which was in accordance with results reported by Dunivin et al. (45), who conducted a global survey of arsenic related genes in soils and identified these phyla as harboring more arsenic resistance genes.

All other elements measured had some influence on the observed bacterial community (Table 1), from key major elements such as sulfur, calcium, and silicon to key trace elements such as iron, manganese, magnesium, zinc, copper, molybdenum, and cadmium, as well as other elements, toxic or not, such as bromine, yttrium, and lead. It should also be noted that while TOC, pH, and conductivity had a significant influence on bacterial community composition (21.8%), the soil elemental composition combined explained most of the variation (>50%). This may serve to highlight the level of complexity of the factors influencing community structure.

**Indicator taxa.** Environmental variables were highly heterogeneous across the landscape, which was reflected by the distribution of alpha diversity and indicator taxa. The indicator species analysis determined abundant OTU-environment associations and identified 163 OTUs that could be considered true specialists in relation to 12 environmental variables. These OTUs were generally associated with high concentrations of the variable in question except for those associated with manganese, niobium, lead, and zirconium, which were representative of low concentrations. As illustrated in the phylogenetic tree (Fig. 5), the diversity of these indicator taxa was high, with numerous representatives of the *Proteobacteria*, *Chloroflexi*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia*. The distribution of some indicator taxa, selected for their reported relationship with the associated variable in the literature, was mapped across the landscape to illustrate the association with the elements' concentrations. For arsenic, *Clostridium* and *Clostridia*-related (*Halanaerobiales*) taxa were mapped, as they have been identified with some role in arsenic cycling (44, 46) and with arsenic resistance genes (45). A *Gemmatimonadetes* and a "*Candidatus* Parcubacteria" (clustered closely with the *Cyanobacteria*) were also mapped, as both have been identified with potential roles in arsenic cycling (46). The distribution of OTUs associated with iron were mapped and included a *Cyanobacteria* (47, 48) and a *Deltaproteobacteria*, a class with known taxa involved in iron cycling (47–49). Finally, the OTUs associated with manganese were also associated with other environmental variables and mainly identified as *Proteobacteria*. A *Deltaproteobacteria* and the only *Chlamydiae* identified were mapped, two classes associated with manganese cycling (48). While this analysis showed the strong associations of some OTUs with the measured environmental parameters, it also illustrated the difficulty of using indicator taxa for monitoring purposes due to the large number of associations identified and the high heterogeneity across the landscape. This was clear when the distribution of key indicator taxa was mapped across the landscape and did not clearly follow the distribution of the environmental variable associated. Furthermore, while indicator taxa may be identified, they do not necessarily participate in the associated element cycle. For instance, these OTUs may benefit from

high concentration of arsenic due to higher tolerance to toxicity and decreased competition, without having any involvement in arsenic cycling.

**Selection and dispersal structure bacterial communities.** The variation partitioning analysis quantified the importance of both selection (deterministic) and dispersal (stochastic) for bacterial community structure. Environmental variables explained 54% of the total variation, corresponding to selection, and 16% were spatially structured, corresponding to the induced spatial dependence. Then, spatial components (trend + dbMEMs) alone explained 10% of the variation, illustrating spatial autocorrelation or dispersal (33). This is the magnitude of influence recorded by Malard et al. (19), who investigated biogeographical patterns across the whole Arctic region, suggesting that the magnitude of influence of dispersal of bacterial community structure may be stable in the Arctic.

More specifically, the distance-decay curve of environmental factors showed that edaphic properties were spatially autocorrelated up to approximately 25 m, although this was the overall spatial autocorrelation, as each variable autocorrelated within different distances. After that distance, environmental variables were independent, and this was illustrated by the weak slope of the linear regression and the overall variability of edaphic properties. In addition, even highly similar environmental conditions could harbor dissimilar bacterial communities, further illustrating the potential role of dispersal and other processes, such as drift or diversification. The distance-decay curve of bacterial communities showed a positive spatial autocorrelation distance at up to 60 m, which was further supported by the Mantel correlogram. For the Arctic region as a whole, an autocorrelation distance within the same order of magnitude, approximating 20 m, was previously identified (19). This limited dispersal range in Arctic soils is in contrast with studies in other regions of the globe. For instance, in a glacier forefield in southern Alaska, this distance was over 600 m (50) while in British soils, it was below 1 km (4). This suggests that Arctic soil bacterial communities disperse only to approximately 60 m and may form rather isolated island communities. Thus, the scale of sampling is important in these landscapes to capture community variability, and therefore, a minimum of 60 m should be maintained between sites to obtain independent samples. Further investigations at other Arctic sites are required to determine whether this applies across the whole Arctic region.

Overall, these results suggest that induced spatial dependence may be an important factor shaping bacterial communities within 25 m; that is, as edaphic properties are very similar, bacterial communities are also similar. Between 25 and 60 m, environmental variability increased, and yet communities remained relatively similar, suggesting that dispersal may be the primary process shaping bacterial communities. Beyond 60 m, the environment was highly heterogeneous, bacterial communities were highly dissimilar, and selection was likely the main process structuring communities. While one process may dominate within each distance category, it is still likely the combination of different processes (selection, dispersal, diversification, and drift) with different magnitudes still driving community assembly (51).

While 63% of the variation (nonadjusted  $R^2 = 81\%$ ) of bacterial community assemblage could be explained, 37% remained unexplained. Many factors, whether biotic or abiotic, could influence bacterial communities. Based on the scale of this study, it is unlikely that most climatic and topographic variables would have much influence on the community structure variation. Instead, other edaphic factors, such as total nitrogen or phosphorus content and clay, silt, and sand content, but also the presence of ice or soil texture may have more impact locally. Furthermore, biotic interactions such as competition and predation within bacterial communities or with other members of the soil biota or higher organisms may have some influence. For instance, grazing is one of the main disturbances to the ecosystem locally, primarily by Svalbard reindeer and the barnacle goose (52). In addition to impacting the vegetation, they trample over the landscape and fertilize it; therefore, grazing can have significant impacts on the ecosystem, and it has been shown to decrease microbial respiration and the

available carbon (53), while animal feces increase the available nitrogen and can increase bacterial abundance (54). Human presence may also have some influence, as the sampling site was close to another scientific research site with open-top chambers. A few cabins were located in the area, and coal mine 7 was still in operation, approximately 1.5 km away and 400 m above the sampling site.

**Conclusion.** In this study, spatial factors accounted for approximately 10% of the variation in community composition at the landscape scale, equivalent to observations across the whole Arctic region, suggesting that while the role and magnitude of other processes involved in community structure may vary, the role of dispersal may be stable globally in the region. Furthermore, the identification of different driving environmental factors at different scales highlights their dependence upon the spatial scales over which they are investigated. Overall, we suggest that induced spatial dependence may shape bacterial communities within 25 m. Between 25 and 60 m, dispersal may be the primary process shaping bacterial communities, and beyond 60 m, selection is likely the main process structuring communities. As dispersal may be limited to 60 m, and while further studies should be conducted, we suggest that soil sampling in the region should be conducted beyond this distance to capture landscape variability when independent samples are being collected. Finally, by mapping the spatial distribution of indicator taxa across the landscape, we showed that strong taxon-environment statistical associations may not actually be reflected in the landscape distribution of these bacterial taxa.

## MATERIALS AND METHODS

**Sampling site.** In July 2017, 44 soil samples were collected in Adventdalen, Svalbard, Norway (Fig. 1A), following the sampling design depicted in Fig. 1B and characterized by 8 north-south transects of 5 samples each. Samples within each transect were approximately 50 m apart, while the distance between transects was approximately 100 m. On transect 6, extra samples were collected 10 m and 1 m apart to investigate smaller-scale patterns (Fig. 1B and C).

Adventdalen is a broad U-shaped valley open to the west, from which the mouth is located approximately 2 km from Longyearbyen and 6 km from Svalbard Airport. Adventdalen was deglaciated about 10 kiloannums before the present (ka BP) (55), and permafrost is estimated to be 100 m thick close to the shore. It is a typical Arctic landscape, in one of the driest areas of Svalbard, with an average of 190 mm of annual precipitation and a mean annual temperature of  $-6^{\circ}\text{C}$  (56). The study site was located approximately 9 km into the valley, 11 km away from Longyearbyen, at  $78.17^{\circ}\text{N}$ ,  $16.02^{\circ}\text{E}$ . The vegetation is primarily dwarf shrub/grass heath, dominated by *Salix* spp., mosses, lichens, and *Graminea* spp. (57) (Fig. 1D). The main disturbances to the site come from grazing, primarily by Svalbard reindeer (*Rangifer tarandus platyrhynchus*) and the barnacle goose (*Branta leucopsis*) (52).

**Sample collection and soil properties.** The coordinates of each site were recorded with a portable global positioning system (GPS) device. At each location, 50 g of soil in the top 15 cm was collected using ethanol-cleaned trowels and Whirl-Pak bags (Nasco, Fort Atkinson, WI, USA). Plant roots and rocks were removed manually in a class II biological safety cabinet (ESCO, Singapore); samples were homogenized and frozen at  $-20^{\circ}\text{C}$  before transportation to the United Kingdom for analyses. pH and conductivity were measured in the laboratory at a 1:5 ratio of freshly thawed soil to water, using a Mettler-Toledo FE20 pH meter (Mettler-Toledo Instruments Co., Shanghai, China) and a CMD500 conductivity meter (WPA, Cambridge, UK). Moisture content was measured gravimetrically on soils after drying at  $150^{\circ}\text{C}$  for 24 h, and total organic content (TOC) was measured gravimetrically by heating previously dried soils to  $550^{\circ}\text{C}$  for 4 h. To analyze the elemental composition of each sample, 5 g of thawed soil was placed in ceramic crucibles and left to dry at  $37^{\circ}\text{C}$  for 5 days. Dried samples were crushed to a fine powder using a mortar and pestle, put in a powder sample cup, placed in the XRF spectrometer (X-Lab2000; Spectro, Kleve, Germany) and analyzed. Resulting concentrations were adjusted using calibrated values and results were expressed in parts per million.

**DNA extraction and amplicon sequencing.** Soil DNA was extracted in duplicate for each sample using the PowerSoil kit (Qiagen, Hilden, Germany) following the manufacturers' protocol. 16S rRNA gene libraries were constructed using the universal primers 515F and 806R (58) to amplify the V4 region. Amplicons were generated using a high-fidelity Accuprime DNA polymerase (Invitrogen, Carlsbad, CA, USA), purified using AMPure magnetic bead capture kit (Agencourt, Beckman Coulter, MA, USA), and quantified using a QuantIT PicoGreen fluorometric kit (Invitrogen). The purified amplicons were then pooled in equimolar concentrations using a SequalPrep plate normalization kit (Invitrogen), and the final concentration of the library was determined using a SYBR green quantitative PCR (qPCR) assay. Libraries were mixed with Illumina-generated PhiX control libraries and our own genomic libraries and denatured using fresh NaOH. The resulting amplicons were sequenced on the Illumina MiSeq V2 (500 cycles).

**Bioinformatics processing.** Raw paired-end reads were subjected to adaptor and primer clipping using Cutadapt (59), resulting in  $71,207 \pm 3,280$  reads per sample. Forward and reverse reads were merged using FLASH (60). Reads with more than 1.5 total expected errors per read and/or read lengths

less than 245 bp were truncated during quality filtration using the Vsearch (61) filtering module. This resulted in  $64,917 \pm 4,291$  high-quality merged reads per sample. Dereplication was performed to identify unique sequences. A two-step chimera detection method was used, first by aligning against ChimeraSlayer Gold database provided with SILVA (62) and second by using the *denovo* detection module. An open-reference operational taxonomic unit (OTU) calling was performed on high-quality trimmed sequences at a 97% similarity level using the USEARCH (63) algorithm for clustering to generate operational taxonomic units (OTUs). This resulted in (85 DNA samples) a total of 5,436,264 reads ( $63,956 \pm 38,865$  reads/sample) assigned against 23,627 OTUs. Unique (chimera-filtered) representative sequences were aligned using the Python Nearest Alignment Space Termination (PyNAST) (64) tool with a relaxed neighbor-joining tree built using FastTree (65). OTUs were assigned taxonomy using the lowest common ancestor (LCA)-based Classification Resources for Environmental Sequence Tags (CREST) package (66), with a minimum classification score of 0.80 against SILVA release 128 as a reference database.

**Statistical analysis.** Alpha diversity (richness and Shannon and Simpson indices) was calculated in QIIME v1.90 (67) on the matrices resulting from multiple rarefactions, with the smallest sample size (22,316 reads) as the maximum depth. Results were collated and averaged to obtain a single representative value for each sample. The OTU table was normalized using cumulative-sum scaling (CSS) (68). The resulting OTU table was input into R for subsequent analyses, and the Bray-Curtis dissimilarity distance was calculated using vegan (69).

To evaluate the environmental component, Pearson's correlation coefficients were calculated using the corplot package (70) to first identify possible correlations between environmental variables. With this many variables, it was a necessary step to avoid misinterpretation of the results (76). Coefficients over  $|0.8|$  indicated strong correlations (Fig. S4), and as such, variables were removed to keep only one representative (76). For example, a high moisture content was correlated with a high TOC content (Pearson's coefficient = 0.88); in this case, moisture was discarded, as it is weather dependent and is expected to be more variable day to day than TOC. Of 48 parameters measured, 35 were independent and considered representative. The distribution of the 35 remaining environmental variables was investigated using the moments package (71) to assess the skewness and kurtosis. Skewness evaluates the degree of distribution shift to one side or another, and a good distribution should equal 0, while kurtosis evaluates the tail distribution and should also be close to 0 to assume normal distribution. Using diagnostic plots, skewness, and kurtosis, the necessary transformations to improve the unimodal distribution of environmental variables were carried out (summarized in Table S2), and collinearity was verified again with Pearson's correlations (Fig. S5). Transformed environment variables were scaled, and a sequential permutational multivariate ANOVA (PERMANOVA) was conducted using the adonis function implemented in vegan with standard 999 permutations to identify environmental variables significantly associated with the Bray-Curtis community dissimilarity.

To evaluate the spatial component, the geographic locations ( $x,y$ ) of the sampling sites were transformed to Cartesian coordinates using the SoDA package (72), and the Euclidean distance was calculated using vegan. Distance-decay curves were produced using linear regressions of the Euclidean distance of the geographic locations against the Bray-Curtis dissimilarity distance and the Euclidean distance of scaled environmental variables.

The presence of a linear trend (a systematic increase or decrease in the OTU data with  $(x,y)$  coordinates) was visualized by the distance-decay curve (Fig. 3A) and tested by redundancy analysis (RDA) and ANOVA, as prescribed by Borcard et al. (33). As a significant linear trend was identified, the OTU table was detrended by linear regression of the  $(x,y)$  coordinates. Distance-based Moran's eigenvector maps (dbMEMs) were constructed with  $(x,y)$  coordinates using the ade4spatial R package (73). The significance of the spatial vectors (dbMEMs) was assessed using the detrended OTU table and tested with ANOVA. Forward selection was conducted to identify significant dbMEM vectors, and the remaining dbMEMs were plotted using RDA.

Variation partitioning analysis (VPA) was used to assess the impact of environmental and spatial factors on community composition (undetrended OTU table) and was conducted using the environmental variables,  $(x,y)$  coordinates (linear trend), and significant dbMEM vectors. Individual fractions were tested using RDA and ANOVA, as prescribed by Borcard et al. (33).

To evaluate spatial autocorrelation, the detrended OTU table and the Euclidean distances of Cartesian coordinates  $(x,y)$  were used to produce a Mantel correlogram with standard 999 permutations using vegan. Semivariograms were also produced using the autoKrige function of the automap package (77) to use for geostatistical analyses. Kriging was conducted using the autoKrig and automapPlot functions in the automap package. Environmental variables and alpha diversity measures were interpolated and mapped across the landscape.

Indicator taxa were determined by the Dufrene-Legendre indicator species analysis (32) to identify OTUs that were specifically associated with different environmental variables. The first step was to define categories for each environmental variable (i.e., high conductivity, medium conductivity, and low conductivity). To identify groups statistically rather than subjectively, an automatic cluster approach was employed using the nbclust package (74), which indicated the ideal number of groups (Table S2). Clusters were created using the kmeans function (Table S2) and used with the multipatt function in the indicspecies package with 999 permutations (32). Indicator taxa with a correlation statistic higher than 0.98 were considered true specialists and used for subsequent analyses. The phylogenetic tree of indicator taxa was built using the representative sequences from the identified indicator taxa using FastTree method (65) and visualized using iTOL (75). Indicator taxon distribution was mapped across the landscape by kriging, as previously described, and Pearson correlations were calculated between the indicator taxa and the environmental variables of interest.

**Data availability.** The sequencing data set is deposited at the European Nucleotide Archive under BioProject accession no. [PRJNA564217](https://www.ebi.ac.uk/bioproject/1564217).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.5 MB.

## ACKNOWLEDGMENTS

This work was supported by a grant from the European Commission's Marie Skłodowska Curie Actions program under project number 675546.

We thank Edwin Sia and UNIS for their participation and support in fieldwork. MiSeq sequencing of the 16S rRNA gene was performed by the NU-OMICS sequencing service (Northumbria University).

L.A.M. and D.A.P. conceived and designed the study and sampling design. L.A.M. carried out the fieldwork and laboratory work. M.Z.A. conducted the bioinformatics processing, and L.A.M. conducted the statistical analysis. L.A.M. drafted the manuscript, and M.Z.A., D.A.P., and C.S.J. revised and approved the final version.

We report no conflict of interests.

## REFERENCES

- Nunan N, Wu K, Young IM, Crawford JW, Ritz K. 2003. Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiol Ecol* 44:203–215. [https://doi.org/10.1016/S0168-6496\(03\)00027-8](https://doi.org/10.1016/S0168-6496(03)00027-8).
- Constancias F, Terrat S, Saby NPA, Horrigue W, Villerd J, Guillemin J-P, Biju-Duval L, Nowak V, Dequiedt S, Ranjard L, Chemidlin Prévost-Bouré N. 2015. Mapping and determinism of soil microbial community distribution across an agricultural landscape. *MicrobiologyOpen* 4:505–517. <https://doi.org/10.1002/mbo3.255>.
- Bru D, Ramette A, Saby N, Dequiedt S, Ranjard L, Jolivet C, Arrouays D, Philippot L. 2011. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J* 5:532–542. <https://doi.org/10.1038/ismej.2010.130>.
- Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. 2011. The bacterial biogeography of British soils. *Environ Microbiol* 13:1642–1654. <https://doi.org/10.1111/j.1462-2920.2011.02480.x>.
- Terrat S, Horrigue W, Dequiedt S, Saby NPA, Lelièvre M, Nowak V, Tripied J, Régnier T, Jolivet C, Arrouays D, Wincker P, Cruaud C, Karimi B, Bispo A, Maron PA, Chemidlin Prévost-Bouré N, Ranjard L. 2017. Mapping and predictive variations of soil bacterial richness across France. *PLoS One* 12: e0186766. <https://doi.org/10.1371/journal.pone.0186766>.
- Zhou J, Deng Y, Shen L, Wen C, Yan Q, Ning D, Qin Y, Xue K, Wu L, He Z, Voordeckers JW, Nostrand JDV, Buzzard V, Michaletz ST, Enquist BJ, Weiser MD, Kaspari M, Waide R, Yang Y, Brown JH. 2016. Temperature mediates continental-scale diversity of microbes in forest soils. *Nat Commun* 7:12083. <https://doi.org/10.1038/ncomms12083>.
- Tedersoo L, Bahram M, Pölme S, Kõljalg U, Yorou NS, Wijesundera R, Villarreal Ruiz L, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson K-H, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L-d, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, et al. 2014. Global diversity and geography of soil fungi. *Science* 346:1256688. <https://doi.org/10.1126/science.1256688>.
- Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. *Science* 359:320–325. <https://doi.org/10.1126/science.aap9516>.
- Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, Huerta-Cepas J, Medema MH, Maltz MR, Mundra S, Olsson PA, Pent M, Pölme S, Sunagawa S, Ryberg M, Tedersoo L, Bork P. 2018. Structure and function of the global topsoil microbiome. *Nature* 560:233–237. <https://doi.org/10.1038/s41586-018-0386-6>.
- Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC. 2011. Drivers of bacterial  $\beta$ -diversity depend on spatial scale. *Proc Natl Acad Sci U S A* 108:7850–7854. <https://doi.org/10.1073/pnas.1016308108>.
- Dao TH. 2014. Landscape-scale geographic variations in microbial biomass and enzyme-labile phosphorus in manure-amended Hapludults. *Biol Fertil Soils* 50:155–167. <https://doi.org/10.1007/s00374-013-0844-0>.
- Palta MM, Ehrenfeld JG, Giménez D, Groffman PM, Subroy V. 2016. Soil texture and water retention as spatial predictors of denitrification in urban wetlands. *Soil Biol Biochem* 101:237–250. <https://doi.org/10.1016/j.soilbio.2016.06.011>.
- Neupane S, Goyer C, Zebarth BJ, Li S, Whitney S. 2019. Soil bacterial communities exhibit systematic spatial variation with landform across a commercial potato field. *Geoderma* 335:112–122. <https://doi.org/10.1016/j.geoderma.2018.08.016>.
- Zak DR, Kling GW. 2006. Microbial community composition and function across an arctic tundra landscape. *Ecology* 87:1659–1670. [https://doi.org/10.1890/0012-9658\(2006\)87\[1659:mccafa\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1659:mccafa]2.0.co;2).
- Chu H, Neufeld JD, Walker VK, Grogan P. 2011. The influence of vegetation type on the dominant soil bacteria, archaea, and fungi in a low Arctic tundra landscape. *Soil Sci Soc Am J* 75:1756–1765. <https://doi.org/10.2136/sssaj2011.0057>.
- Bottos EM, Kennedy DW, Romero EB, Fansler SJ, Brown JM, Bramer LM, Chu RK, Tfaily MM, Jansson JK, Stegen JC. 2018. Dispersal limitation and thermodynamic constraints govern spatial structure of permafrost microbial communities. *FEMS Microbiol Ecol* 94:fy110. <https://doi.org/10.1093/femsec/fiy110>.
- Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* 12:2998–3006. <https://doi.org/10.1111/j.1462-2920.2010.02277.x>.
- Siciliano SD, Palmer AS, Winsley T, Lamb E, Bissett A, Brown MV, van Dorst J, Ji M, Ferrari BC, Grogan P, Chu H, Snape I. 2014. Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biol Biochem* 78:10–20. <https://doi.org/10.1016/j.soilbio.2014.07.005>.
- Malard LA, Anwar MZ, Jacobsen CS, Pearce DA. 2019. Biogeographical patterns in soil bacterial communities across the Arctic region. *FEMS Microbiol Ecol* 95:fi28. <https://doi.org/10.1093/femsec/fiz128>.
- Dimitriu PA, Grayston SJ. 2010. Relationship between soil properties and patterns of bacterial  $\beta$ -diversity across reclaimed and natural boreal forest soils. *Microb Ecol* 59:563–573. <https://doi.org/10.1007/s00248-009-9590-0>.
- Docherty KM, Borton HM, Espinosa N, Gebhardt M, Gil-Loaiza J, Gutknecht JLM, Maes PW, Mott BM, Parnell JJ, Purdy G, Rodrigues PAP, Stanish LF, Walsler ON, Gallery RE. 2015. Key edaphic properties largely explain temporal and geographic variation in soil microbial communities

- across four biomes. *PLoS One* 10:e0135352. <https://doi.org/10.1371/journal.pone.0135352>.
22. Katsalirou E, Deng S, Nofziger DL, Gerakis A, Fuhlendorf SD. 2010. Spatial structure of microbial biomass and activity in prairie soil ecosystems. *Eur J Soil Biol* 46:181–189. <https://doi.org/10.1016/j.ejsobi.2010.04.005>.
  23. Dequiedt S, Saby N, Lelievre M, Jolivet C, Thioulouse J, Toutain B, Arrouays D, Bispo A, Lemanceau P, Ranjard L. 2011. Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Global Ecol Biogeogr* 20:641–652. <https://doi.org/10.1111/j.1466-8238.2010.00628.x>.
  24. Bahram M, Kohout P, Anslan S, Harend H, Abarenkov K, Tedersoo L. 2016. Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME J* 10:885–896. <https://doi.org/10.1038/ismej.2015.164>.
  25. Baker KL, Langenheder S, Nicol GW, Ricketts D, Killham K, Campbell CD, Prosser JL. 2009. Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol Biochem* 41:2292–2298. <https://doi.org/10.1016/j.soilbio.2009.08.010>.
  26. Franklin RB, Mills AL. 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol Ecol* 44:335–346. [https://doi.org/10.1016/S0168-6496\(03\)00074-6](https://doi.org/10.1016/S0168-6496(03)00074-6).
  27. Taberlet P, Prud'Homme SM, Campione E, Roy J, Miquel C, Shehzad W, Gielly L, Rioux D, Choler P, Clément J-C, Melodelima C, Pompanon F, Coissac E. 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol Ecol* 21:1816–1820. <https://doi.org/10.1111/j.1365-294X.2011.05317.x>.
  28. Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficaretola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kausarud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P. 2019. DNA metabarcoding—need for robust experimental designs to draw sound ecological conclusions. *Mol Ecol* 28:1857–1862. <https://doi.org/10.1111/mec.15060>.
  29. Simonin M, Voss KA, Hassett BA, Rocca JD, Wang SY, Bier RL, Violin CR, Wright JP, Bernhardt ES. 2019. In search of microbial indicator taxa: shifts in stream bacterial communities along an urbanization gradient. *Environ Microbiol* 21:3653–3668. <https://doi.org/10.1111/1462-2920.14694>.
  30. Yang Y, Li S, Gao Y, Chen Y, Zhan A. 2019. Environment-driven geographical distribution of bacterial communities and identification of indicator taxa in Songhua River. *Ecol Indicators* 101:62–70. <https://doi.org/10.1016/j.ecolind.2018.12.047>.
  31. Hermans SM, Buckley HL, Case BS, Curran-Courne F, Taylor M, Lear G. 2017. Bacteria as emerging indicators of soil condition. *Appl Environ Microbiol* 83:e02826-16. <https://doi.org/10.1128/AEM.02826-16>.
  32. Cáceres MD, Legendre P. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* 90:3566–3574. <https://doi.org/10.1890/08-1823.1>.
  33. Borcard D, Gillet F, Legendre P. 2018. Numerical ecology with R. Springer, New York, NY.
  34. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* 103:626–631. <https://doi.org/10.1073/pnas.0507535103>.
  35. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* 4:1340–1351. <https://doi.org/10.1038/ismej.2010.58>.
  36. Siles JA, Margesin R. 2016. Abundance and diversity of bacterial, archaeal, and fungal communities along an altitudinal gradient in alpine forest soils: what are the driving factors? *Microb Ecol* 72:207–220. <https://doi.org/10.1007/s00248-016-0748-2>.
  37. Wojcik R, Donhauser J, Frey B, Holm S, Holland A, Anesio AM, Pearce DA, Malard L, Wagner D, Benning LG. 2018. Linkages between geochemistry and microbiology in a proglacial terrain in the High Arctic. *Ann Glaciol* 59:95–110. <https://doi.org/10.1017/aog.2019.1>.
  38. Malard LA, Pearce DA. 2018. Microbial diversity and biogeography in Arctic soils. *Environ Microbiol Rep* 10:611–625. <https://doi.org/10.1111/1758-2229.12680>.
  39. Post WM, Emanuel WR, Zinke PJ, Stangenberger AG. 1982. Soil carbon pools and world life zones. *Nature* 298:156–159. <https://doi.org/10.1038/298156a0>.
  40. Callaghan TV, Johansson M, Brown RD, Groisman PY, Labba N, Radionov V, Barry RG, Bulygina ON, Essery RLH, Frolov DM, Golubev VN, Grenfell TC, Petrushina MN, Razuvaev VN, Robinson DA, Romanov P, Shindell D, Shmakin AB, Sokratov SA, Warren S, Yang D. 2011. The changing face of Arctic snow cover: a synthesis of observed and projected changes. *Ambio* 40:17–31. <https://doi.org/10.1007/s13280-011-0212-y>.
  41. Bonnaventure PP, Lamoureux SF. 2013. The active layer: a conceptual review of monitoring, modelling techniques and changes in a warming climate. *Progress in Physical Geography* 37:352–376. <https://doi.org/10.1177/0309133313478314>.
  42. Piña RG, Cervantes C. 1996. Microbial interactions with aluminium. *Bio-metals* 9:311–316. <https://doi.org/10.1007/BF00817932>.
  43. Barabasz W, Albinska D, Jaskowska M, Lipiec J. 2002. Ecotoxicology of aluminium. *Polish J Environ Studies* 11:199–204.
  44. Oremland RS, Stolz JF. 2003. The ecology of arsenic. *Science* 300:939–944. <https://doi.org/10.1126/science.1081903>.
  45. Dunivin TK, Yeh SY, Shade A. 2019. A global survey of arsenic-related genes in soil microbiomes. *BMC Biol* 17:45. <https://doi.org/10.1186/s12915-019-0661-5>.
  46. Jia Y, Huang H, Zhong M, Wang F-H, Zhang L-M, Zhu Y-G. 2013. Microbial arsenic methylation in soil and rice rhizosphere. *Environ Sci Technol* 47:3141–3148. <https://doi.org/10.1021/es303649v>.
  47. Emerson D, Scott JJ, Benes J, Bowden WB. 2015. Microbial iron oxidation in the arctic tundra and its implications for biogeochemical cycling. *Appl Environ Microbiol* 81:8066–8075. <https://doi.org/10.1128/AEM.02832-15>.
  48. Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. 2017. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geosci* 10:629–636. <https://doi.org/10.1038/ngeo3006>.
  49. Lipson DA, Raab TK, Parker M, Kelley ST, Brislaw CJ, Jansson J. 2015. Changes in microbial communities along redox gradients in polygonized Arctic wet tundra soils. *Environ Microbiol Rep* 7:649–657. <https://doi.org/10.1111/1758-2229.12301>.
  50. Darcy JL, King AJ, Gendron E, Schmidt SK. 2017. Spatial autocorrelation of microbial communities atop a debris-covered glacier is evidence of a supraglacial chronosequence. *FEMS Microbiol Ecol* 93:fix095. <https://doi.org/10.1093/femsec/fix095>.
  51. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach A-L, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 4:102–112. <https://doi.org/10.1038/nrmicro1341>.
  52. Strebler D, Elberling B, Morgner E, Knicker HE, Cooper EJ. 2010. Cold-season soil respiration in response to grazing and warming in High-Arctic Svalbard. *Polar Res* 29:46–57. <https://doi.org/10.1111/j.1751-8369.2010.00154.x>.
  53. Stark S, Grellmann D. 2002. Soil microbial responses to herbivory in an arctic tundra heath at two levels of nutrient availability. *Ecology* 83:2736–2744. [https://doi.org/10.1890/0012-9658\(2002\)083\[2736:SMRTHI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2736:SMRTHI]2.0.CO;2).
  54. Sørensen LI, Mikola J, Kytöviita M-M, Olofsson J. 2009. Trampling and spatial heterogeneity explain decomposer abundances in a sub-arctic grassland subjected to simulated reindeer grazing. *Ecosystems* 12:830–842. <https://doi.org/10.1007/s10021-009-9260-6>.
  55. Mangerud J, Bolstad M, Elgersma A, Helliksen D, Landvik JY, Lønne I, Lycke AK, Salvigsen O, Sandahl T, Svendsen JI. 1992. The last glacial maximum on Spitsbergen, Svalbard. *Quat Res* 38:1–31. [https://doi.org/10.1016/0033-5894\(92\)90027-G](https://doi.org/10.1016/0033-5894(92)90027-G).
  56. Christiansen HH. 2005. Thermal regime of ice-wedge cracking in Adventdalen. *Permafrost Periglacial Process* 16:87–98. <https://doi.org/10.1002/ppp.523>.
  57. Mora C, Vieira G, Pina P, Lousada M, Christiansen HH. 2015. Land cover classification using high-resolution aerial photography in Adventdalen, Svalbard. *Geografiska Annal Ser A Phys Geogr* 97:473–488. <https://doi.org/10.1111/geoa.12088>.
  58. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R. 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* 1:e00009-15. <https://doi.org/10.1128/mSystems.00009-15>.
  59. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
  60. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/bt507>.
  61. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile

- open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>.
62. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35:7188–7196. <https://doi.org/10.1093/nar/gkm864>.
  63. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
  64. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNASt: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. <https://doi.org/10.1093/bioinformatics/btp636>.
  65. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
  66. Lanzén A, Jørgensen SL, Huson DH, Gorfer M, Grindhaug SH, Jonassen I, Øvreås L, Urich T. 2012. CREST—classification resources for environmental sequence tags. *PLoS One* 7:e49334. <https://doi.org/10.1371/journal.pone.0049334>.
  67. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
  68. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* 10:1200–1202. <https://doi.org/10.1038/nmeth.2658>.
  69. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
  70. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. 2017. Package ‘corrplot’. *Statisticians* 56:316–324.
  71. Komsta L, Novomestky F. 2015. Moments, cumulants, skewness, kurtosis and related tests. R package version 14.
  72. Chambers J. 2008. Software for data analysis: programming with R. Springer Science & Business Media, Berlin, Germany.
  73. Dray S, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G, Legendre P, Madi N, Wagner H. 2017. Adespatial: multivariate multi-scale spatial analysis. R package version 0.0–9.
  74. Charrad M, Ghazzali N, Boiteau V, Niknafs A, Charrad MM. 2014. Package ‘nbclust’. *J Stat Soft* 61:1–36. <https://doi.org/10.18637/jss.v061.i06>.
  75. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
  76. Katz MH. 2011. *Multivariable analysis: a practical guide for clinicians and public health researchers*. Cambridge University Press, Cambridge, United Kingdom.
  77. Hiemstra P. 2013. Package ‘automap.’ <https://cran.r-project.org/web/packages/automap/automap.pdf>.