Title: Biogeographical patterns in soil bacterial communities across the Arctic region

Authors:

Lucie A Malard[1, *]

Muhammad Zohaib Anwar[2]

Carsten S Jacobsen[2]

David A Pearce[1,3]

Author affiliations:

[1]Faculty of Health and Life Sciences

Northumbria University

Newcastle-upon-Tyne NE1 8ST

United Kingdom

[2]Department of Environmental Sciences

Aarhus University

4000 Roskilde

Denmark

[3]British Antarctic Survey

High Cross Madingley Road

Cambridge

CB3 0ET

United Kingdom

*Corresponding Author:

Lucie Malard: lucie.malard@northumbria.ac.uk

Proposed keywords: 16S rRNA, Arctic soil, microbial diversity, indicator species, core microbiome, biogeography

LAM and DAP conceived and designed the study and sampling strategy. LAM carried the fieldwork and laboratory work. MZA conducted bioinformatics processing and statistical analysis. LAM drafted the manuscript and MZA, DAP and CSJ revised and approved the final version.

Abstract

The considerable microbial diversity of soils, their variety and key role in biogeochemical cycling has led to growing interest in their global distribution and the impact that environmental change might have at the regional level. In the broadest study of Arctic soil bacterial communities to date, we used high-throughput DNA sequencing to investigate the bacterial diversity from 200 independent Arctic soil samples from 43 sites. We quantified the impact of spatial and environmental factors on bacterial community structure using variation partitioning analysis, illustrating a non-random distribution across the region. pH was confirmed as the key environmental driver structuring Arctic soil bacterial communities, while total organic carbon, moisture and conductivity were shown to have little effect. Specialist taxa were more abundant in acidic and alkaline soils while generalist taxa were more abundant in acidoneutral soils. Of 48,147 bacterial taxa, a core microbiome composed of only 13 taxa that were ubiquitously distributed and present within 95% of samples was identified, illustrating the high potential for endemism in the region. Overall, our results demonstrate the importance of spatial and edaphic factors on the structure of Arctic soil bacterial communities.

Introduction

Frozen soils in the Arctic region store over 1500 Pg of carbon (Koven et al., 2011; Mackelprang et al., 2011) and as Arctic warming is exacerbated and permafrost thaw accelerates, the depth of the active layer is increasing. As previously frozen carbon becomes available, it is expected that microbial activity will increase, which may lead to increased atmospheric release rates of climate active gases such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (Ma et al., 2007; Mackelprang et al., 2011). Carbon-climate feedback studies of permafrost affected regions use temperature, soil moisture and precipitation as the main drivers of decomposition rates (Koven et al., 2011; Schuur et al., 2015). While models are useful to gain a global understanding of the impact of climate change on permafrost thaw and greenhouse gas release, the accuracy of results obtained is highly variable when compared with data collected in the field or laboratory (Schuur et al., 2015) due to empirical and modelling uncertainties which still need to be addressed (Bradford et al., 2016).

As microorganisms drive biogeochemical cycling and participate in the uptake and release of CO_2, CH_4 and N_2O, microbial data should be incorporated in climate models. Current models use soil properties to model changes in fluxes, without considering microbial communities and the changes in community composition induced by climate change (Bardgett et al., 2008; Nazaries et al., 2013). Adding microbial information into models will improve their predictions; however, detailed microbial data is still required, with a focus on the microbial community, diversity, function, stability and long-term changes in these communities (Graham et al., 2012; Nazaries et al., 2013). While global surveys of microbial diversity have already been conducted (Tedersoo et al., 2014; Delgado-Baquerizo et al., 2018), the number of Arctic samples is restricted (Malard and Pearce, 2018) and therefore, microbial data is still lacking for most permafrost-affected regions.

To accurately model spatial distribution patterns of microbes and incorporate microbial data from permafrost-affected regions in climate models in the near future, the first step is to produce a baseline database of microbial diversity with respect to biogeographical distribution. Biogeography is the study of biodiversity across space and time, it gives insights into ecological mechanisms such as speciation, extinction, dispersal and species interactions (Martiny et al., 2006; Fierer, 2008), improving our understanding of community assembly. Theoretically, distant and isolated habitats are expected to present high endemicity as a consequence of intrinsic dispersal limitations and environmental filtering (Mittelbach and Schemske, 2015; Kleinteich et al., 2017; Bahram et al., 2018). Thus, isolated, pristine ecosystems with limited human presence, such as the Arctic region, should harbour endemic communities. However, microbial communities may be less constrained by geographical barriers and thus, have long been considered ubiquitous (Finlay, 2002; O'Malley, 2007). Yet, recent studies have uncovered patterns of microbial biogeography on global scales (Fierer and Jackson, 2006; Lauber et al., 2009; Tedersoo et al., 2014; Henschel et al., 2015; Bahram et al., 2018; Delgado-Baquerizo et al., 2018). They demonstrated that microbial distribution is non-random, generally influenced by biotic and abiotic factors such as plant cover, soil pH, C/N ratio or precipitation, and that communities vary from one region to another. However, whether these patterns apply to Arctic microbial communities is still

subject to debate as these studies tend to have a low number of Arctic samples and Arctic studies have generally focused on small scales patterns in restricted areas, as shown by Metcalfe et al. (2018) highlighting the focuses on Abisko, Sweden and Toolik lake, Alaska. Large scale studies are therefore required to improve our understanding not only of the spatial distribution of microorganisms but also on the processes of community assembly. Producing a baseline database of diversity in the region will enable the accurate modelling of spatial distribution, linking with patterns of functional processes to evaluate the potential consequences of environmental change on ecosystem properties.

In this study, we conducted a Pan-Arctic survey of bacterial communities in Arctic soils to provide a baseline of Arctic bacterial diversity. We addressed the following questions: (i) Are bacterial taxa ubiquitously distributed or are there biogeographical patterns ? (ii) What are the key factors influencing bacterial community structure ? (iii) Are there taxa closely associated with the key edaphic properties and what can we learn from them? (iv) If bacterial taxa are not ubiquitous, is there a core microbiome?

Methods

Sample collection

Soil samples were collected across the Arctic region between April 2017 and September 2017, the GPS coordinates of each site were recorded with a portable GPS [Fig. 1]. All sites were generally remote with limited human presence; however, in Iceland and Norway where tourist activities are frequent, we aimed to sample at least 500 m away from human-related activities. At each location, 3 to 5 soil samples were collected within 100 m^2 under the most common vegetation, for a total of 200 independent Arctic samples across 43 sites, with the aim to cover the broadest geographical area possible. Despite the extent, it should be noted that large areas of Canada and Russia remain uncharacterized and under-represented, highlighting areas where sampling should be prioritized in the future. Approximately 150 g of soil per sample was collected from the top 15 cm, avoiding plant roots and rocks and using an ethanol-cleaned shovel and Whirl-Pak bags (Nasco, WI, USA). Remaining plant roots and rocks were removed manually in a class II microbiological safety cabinet; samples were

110 homogenized by manual mixing and frozen at -20 °C before transportation to the United Kingdom.
111 Samples were conserved at – 20 °C until analysed.

112 ***Soil properties***

113 We focused our investigation on environmental variables previously identified as influencing bacterial
114 communities and included pH, conductivity, moisture and organic carbon. Moisture content was
115 measured gravimetrically for each soil sample after drying at 150 °C for 24 h and total organic content
116 (TOC) was measured gravimetrically by heating previously dried soils to 550 °C for 4 h. pH and
117 conductivity were measured in the laboratory in a 1:5 freshly thawed soil to water ratio, using a
118 Mettler-Toledo FE20 pH meter (Mettler-Toledo Instruments co., Shanghai, China) and a CMD500
119 conductivity meter (WPA, Cambridge, UK).

120 ***DNA extraction***

121 Soil DNA was extracted for each sample using the PowerSoil kit (Qiagen, Hilden, Germany), for a total
122 of 400 DNA extracts. Each sample was PCR amplified using the universal primers 515F-806R, as per
123 the Schloss lab standard operating Procedure (Kozich et al., 2013) and the Earth Microbiome Project
124 (Thompson et al., 2017), under the following conditions: initial denaturation at 95°C for 2 min then 30
125 cycles of 20 s denaturation at 95°C; primer annealing at 55°C for 15 s; elongation at 72°C for 5 min
126 then a final elongation at 72°C for 10 min. Negative controls, DNA extraction kit controls and
127 ZymoBIOMICS mock communities (Zymo Research, Irvine, CA, USA) were included alongside the soil
128 DNA and sequenced. PCR amplicons were cleaned and normalized using SequalPrep Plate
129 Normalization Kit (Invitrogen, Carlsbad, CA, USA) and combined into four pools. Each pool was
130 quantified using fragment size determined by BioAnalyzer hsDNA assay (Agilent technologies, Santa
131 Clara, CA, USA) and concentration by Qubit hsDNA kit (Invitrogen). The library was supplemented with
132 5% PhiX and loaded on an Illumina MiSeq V2 500 cycles cartridge.

133

134 ***Illumina Sequencing and Data Processing***

135 Raw amplicon sequences were demultiplexed with the associated barcodes. Cutadapt (Martin, 2011)
136 was used for adaptor and primer clipping. Forward and reverse reads longer than 240 bp were merged
137 (98 % ± 0.8 % / sample) using FLASH (fast length adjustment of short reads) (Magoč and Salzberg, 2011)
138 for a total of 20 million reads (~50,000 ± 30,000 reads/sample) initially. Vsearch (Rognes et al., 2016)
139 was used for downstream analyses. Quality filtering was carried out with an expected error < 1.5.
140 Dereplication was performed to identify unique sequences. A two-step chimera detection method was
141 used, first by aligning against ChimeraSlayer Gold database provided with SILVA (Pruesse et al., 2007),
142 second by using the *de novo* detection module in vsearch. *De novo* operational taxonomic unit (OTU)
143 calling was performed on high-quality trimmed sequences at 97% similarity level using the USEARCH
144 (Edgar, 2010) algorithm for clustering implemented in vsearch to generate operational taxonomical
145 units (OTUs). Unique chimera filtered sequences were aligned using the Python Nearest Alignment
146 Space Termination (PyNAST) (Caporaso et al., 2009) tool with a relaxed neighbour-joining tree built
147 using FastTree (Price et al., 2010). The taxonomy was determined using the Classification Resources
148 for Environmental Sequence Tags (CREST) (Lanzén et al., 2012) classifier with a confidence threshold
149 of 0.80 against SILVA release 128 as a reference database.

150 Samples less than at least 2000 reads/sample were filtered from the OTU table in order to have
151 sufficient reads as described in Caporaso et al. (2010a). After filtering, 386 samples were used for the
152 statistical analyses, corresponding to 386 DNA extracts from 200 independent samples and ~19.5
153 million reads (50,609 ± 26,700 reads/sample) assigned against 48,147 OTUs.

154 ***Data Availability***

155 The dataset is deposited at European Nucleotide Archive / SRA under the accession number
156 PRJEB29109.

157

158 ***Statistical Analysis***

159 All statistical analyses were performed with a combination of QIIME1 V 1.90 (Caporaso et al., 2010b)
160 and the R environment (Team, 2013). Multiple rarefactions were performed in QIIME with the
161 smallest sample size as maximum depth. Alpha diversity (richness, Shannon and Simpson indices)
162 were calculated in QIIME on the matrices resulting from the multiple rarefactions. The differences in
163 alpha diversity indices were tested in QIIME using non-parametric (Monte-Carlo) tests across different
164 pH categories with a Bonferroni correction.
165 The non-rarefied OTU-table was normalized using cumulative-sum scaling (CSS) in QIIME (Paulson et
166 al., 2013). The resulting OTU-table was input into R for subsequent analyses. Using the phyloseq
167 package (McMurdie and Holmes, 2013), we calculated the Bray-Curtis dissimilarity distance and
168 visualized the ordination using principal coordinate analysis (PCoA). Using the cmdscale, envfit and
169 ordisurf functions in the vegan package (Dixon, 2003), we calculated and visualized the significant
170 correlations between edaphic properties and bacterial community dissimilarity in relation to the pH
171 gradient. Permutational multivariate analyses of variance (PerMANOVA) was conducted using the
172 adonis function with 999 permutations. Pearson's correlation with 999 permutations was used to
173 evaluate the collinearity of environmental variables.

174 To evaluate the spatial component, the geographic locations of the sampling sites were transformed into
175 cartesian coordinates using the SoDA package (Chambers, 2008) and the euclidean distance was
176 calculated using the vegan package. The presence of a linear trend was tested by redundancy
177 discriminant analysis and ANOVA as prescribed in Borcard et al. (2018). A significant linear trend was
178 identified, violating the second-order stationarity assumption where the mean of the variable and its
179 spatial covariance are the same over the study area and its variance is finite. In other words, spatial
180 correlation coefficients cannot be tested for significance if an overall trend is present in the data
181 (Franklin and Mills, 2007; Borcard et al., 2018) and therefore, the data was detrended by linear
182 regression of the x,y coordinates. To carry meaningful spatial analysis, constructing spatial variables

183 representing spatial structures at all relevant scales is necessary (Borcard et al., 2018). To do so, we
184 used distance-based Moran's Eigenvector Maps (dbMEM) on detrended data and x,y coordinates using
185 the adespatial R package (Dray et al., 2017). Significance of the spatial vectors was assessed using
186 ANOVA. Forward selection was carried out to identify significant dbMEM vectors.

187 Variation partitioning analysis (VPA) was used to assess the impact of environmental and spatial
188 factors on Bray-Curtis dissimilarity. The linear trend was considered to represent a source of variation
189 like any other and therefore, we performed variation partitioning analysis using the environmental
190 variables, x,y coordinates (trend) and significant dbMEM vectors.

191 Generalist and specialist taxa were identified by selecting taxa present in over 95 % of samples within
192 each unique and overlapped pH category. Indicator taxa were determined by the Dufrene-Legendre
193 indicator species analysis using the multipatt function (Cáceres and Legendre, 2009) to identify abundant
194 OTUs that were specifically associated with the different pH ranges. The phylogenetic tree of indicator
195 species was built using the aligned sequences from the identified indicator OTUs and input into FastTree.
196 The tree was visualized using the Interactive Tree of Life (iTOL) online tool (Letunic and Bork, 2016).

197 Results

198 *Bacterial community composition*

199 We identified 48,147 bacterial taxa, mainly classified as Proteobacteria (20 %) and Planctomycetes
200 (15%). Acidobacteria represented 13% while Actinobacteria, Bacteroidetes, Chloroflexi and
201 Verrucomicrobia each accounted for 4 – 6 % of the community. Of these 48,147 OTUs, only 135 had
202 overall abundances over 0.1% across all 386 samples and were defined as abundant taxa. They
203 accounted for 32% of all the reads, illustrating the dominance of a few taxa over the rest of the
204 community. Within these abundant taxa, Acidobacteria dominated the sub-community at
205 approximately 32% with Blastocatellia (12.8%) and Acidobacteria subgroup6 (7.2%) as abundant
206 classes. Alphaproteobacteria (10.6%) and Betaproteobacteria (6.6%) were the most commonly
207 identified Proteobacteria (21% overall) and Verrucomicrobia was largely represented by

208 Spartobacteria (17.2%). Actinobacteria (10.7%), Chloroflexi (7%) and Bacteroidetes (3.5%) were also
209 among the abundant phyla classified.

210 ***Spatial and environmental factors influencing bacterial communities***

211 Distance-decay curves were used to evaluate the influence of spatial factors on bacterial community
212 structure. Figure 2A illustrate the increase in community dissimilarity with increasing geographic
213 distance (linear regression: $R^2 = 0.108$, $p < 2.2 \times 10^{-16}$). The best-fitted power model ($R^2 = 0.221$, p =
214 0.005) suggested the presence of spatial autocorrelation, also qualified as dispersal limitation, within
215 the first 500 m. For a more accurate estimate of dispersal limitation, we plotted the same graph within
216 0 and 500 m [Fig. 2B] and observed spatial autocorrelation approximately up to 20 m.
217 To determine the influence of environmental factors on the bacterial community structure, we used the
218 adonis function with 999 permutations. Results indicated that pH explained 18.7 % of the variance in
219 bacterial community composition overall [Table 1]. In comparison, conductivity, moisture and total
220 organic carbon (TOC) explained almost 10 % of the variance all together. The measured environmental
221 parameters explained 29 % of the variance in the bacterial sub-community, leaving 71 % unexplained.
222 To quantify the influence of edaphic properties and spatial factors, we conducted variation
223 partitioning on the bacterial community. We built spatial vectors using distance-based Moran's
224 Eigenvector Maps (dbMEM) on x,y geographic coordinates. Three vectors (MEMs) were produced;
225 after forward selection, two significant MEMs were identified for the total community and used in
226 further analyses. The variation partitioning analysis differentiated the effect of environmental
227 factors, linear trend and spatial vectors on Bray-Curtis dissimilarity [Fig.3]. Results indicated that the
228 environmental factors (X1 = [a]+[d]+[g]+[f]) explained 30.7 % of the variance in the community. These
229 results are equivalent to the finding of the adonis test [Table 1], confirming the success of the
230 variation partitioning analysis. Non-spatially structured environmental variables (fraction [a])
231 explained 15 % of the variation while spatial factors only (fractions [b], [e] and [c]) explained
232 approximately 10 % of variation. Of fraction X1 (environmental factors), 15 % represented spatially

233 structured environmental variables, also called induced spatial dependence, where the spatial
234 structure of these environmental variables induced a similar spatial structure in the response data
235 [detailed results in Table S1]. In total, 63.3 % of the total bacterial community dissimilarity could not
236 be explained by environmental and spatial factors after R^2 adjustment.

237 ***pH as a key environmental factor***

238 Of the environmental variables we measured, we identified pH as the primary factor structuring Arctic
239 soil bacterial communities. The Pearson correlations indicated that pH did not have strong collinearity
240 (coefficient over |0.8|) with any other variable measured [Table S2]. The clustering of samples by pH
241 range was clearly observed on the principal coordinate analysis (PCoA) of bacterial communities [Fig.
242 4A]. The envfit analysis confirmed that pH was the key measured factor driving the ordination,
243 although TOC, conductivity and moisture all had some influence. Using ordisurf, we visualised the PCoA
244 ordination of the bacterial community in relation to the pH gradient [Fig. S1]. The Bray-Curtis
245 dissimilarity heatmap and dendrogram [Fig. 4B] identified three main clusters illustrating community
246 differences. The first cluster was composed of acidic samples in a pH gradient from Norwegian soils at
247 pH = 4.07 (± 0.35) to samples from Alaska (pH = 4.64 ± 0.41) and West Greenland (pH = 4.90 ± 0.85).
248 The second cluster included the lower acidoneutral range of samples from East Greenland (pH = 5.96
249 ± 0.69), Svalbard (pH = 5.65 ± 0.53) and Iceland (pH = 5.84 ± 0.46). Finally, the last cluster included the
250 higher range of acidoneutral and alkaline soils from Russia (pH = 6.19 ± 0.22) and Canada (pH = 7.94 ±
251 0.67). The full spectrum of soil pH was covered from pH = 3.5 to pH = 9.0 and geographical locations
252 often had samples from more than one pH category [Fig. S2]. As pH was the primary factor we identified
253 influencing bacterial community structure, we defined three pH categories to focus on: acidic,
254 acidoneutral and alkaline.

255

256

257

258 ***Bacterial diversity by pH category***

259 We compared the observed richness and Shannon diversity index by pH range [Fig. 5A, B] using a
260 Monte-Carlo test to assess differences. Alpha-diversity was significantly lower in acidic samples than
261 in acidoneutral and alkaline soils [Table S3]

262 In terms of community composition, we differentiated the abundant community, composed of only
263 135 OTUs, from the total community composed of 48,147 OTUs and observed the changes in relative
264 abundance by pH range. In the total community, Acidobacteria, Armatimonadetes, Planctomycetes and
265 Verrucomicrobia decreased with increasing pH. On the other hand, Bacteroidetes, Chloroflexi and
266 Gemmatimonadetes increased in alkaline soils; Proteobacteria remained stable throughout. In the
267 abundant community, Acidobacteria, Bacteroidetes, Gemmatimonadetes and SPAM increased with
268 increasing pH while Actinobacteria, Planctomycetes and Verrucomicrobia remained stable [Fig. 6].

269 ***Generalist vs specialist taxa***

270 The differentiation of specialist from generalist taxa associated with pH range was conducted by
271 considering all taxa identified in this study and present in a minimum 95% of all samples from each pH
272 category [Fig. 7]. 125 acidic specialist OTUs were identified and unique to acidic soil samples. Of these
273 125 unique OTUs, most belonged to the Acidobacteria (27%), Verrucomicrobia (20%), Actinobacteria
274 (14%), Planctomycetes (14%) and Proteobacteria (14%). At the class level, Acidobacteria group 1
275 dominated at 14%, while the rest of the identified classes had balanced relative abundances oscillating
276 between 5 and 8%. Shared OTUs, or generalists were present in low numbers in acidic soils. Only 12
277 OTUs were shared with acidoneutral soils only, which were dominated by Alphaproteobacteria (65%),
278 mainly Rhizobiales, while 3 OTUs were shared with alkaline only, belonging to the Spartobacteria
279 (Verrucomicrobia), Phycisphaerae (Planctomycetes) and Chitinophagia (Bacteroidetes) classes. In
280 comparison, acidoneutral soils had 76 shared taxa with alkaline soils, against 12 shared OTUs with
281 acidic. Taxa shared with alkaline soils were mainly classified as Blastocatellia (Acidobacteria),
282 Spartobacteria and Acidobacteria subgroup 6. Taxa exclusively found in acidoneutral soils belonged

283 mainly to Actinobacteria (20%), Verrucomicrobia (20%) and Acidobacteria (19%). At the class level,
284 unique taxa living in acidoneutral soils were dominated by Spartobacteria, Alphaproteobacteria and
285 Holophagae. Alkaline soils presented a combination of both, a high number of shared (79 OTUs in
286 total) and 125 exclusive taxa. Alkaline unique taxa were mostly composed of Acidobacteria (36%) and
287 Proteobacteria (22%). From these unique taxa, the Acidobacteria subgroup 6 (20%), Blastocatellia
288 (12%) and Alphaproteobacteria (12%) dominated the community.

289 ***Indicator species***

290 We conducted the indicator species analysis on the abundant taxa only as they may be more useful for
291 monitoring purposes. We identified 17 unique taxa-habitat associations. In acidic soils, 10 indicator
292 species, mainly Acidobacteria group 1 and 2 and Ca. Methylacidiphilum were identified. The 6 OTUs
293 associated with Acidoneutral soils were mainly Acidobacteria group 4 (Blastocatellia) and
294 Spartobacteria (Verrucomicrobia). Finally, only 1 OTU was identified as an indicator species for alkaline
295 soils and belonged to the Holophagae (Acidobacteria). We also conducted the indicator species analysis
296 of abundant taxa to combine pH ranges and identified 84 OTUs. 21 were identified as indicator species
297 of acidic and acidoneutral soils combined, and mainly belonged to the Verrucomicrobia. Only 2 OTUs
298 were associated with acidic and alkaline soils, Acidothermus (Actinobacteria) and Candidatus
299 Xiphinematobacter (Verrucomicrobia), further illustrating the low overlap of taxa between these
300 ecosystems and inferring the large ecosystem differences. Finally, 61 OTUs were associated with
301 acidoneutral and alkaline soils, mainly belonging to the Acidobacteria and Proteobacteria phyla (see
302 table S3 for the detailed list). The phylogenetic tree [Fig.S3] illustrates the classification of the
303 identified abundant indicator taxa with the pH range associated. Although displaying diversity, they
304 were mainly classified as Acidobacteria and Verrucomicrobia, with the vast majority of taxa
305 associated with acidoneutral and alkaline soils.

306

307

308 ***The core microbiome***

309 While some taxa displayed biogeographical patterns with non-random distribution across space and
310 environmental gradients, others were cosmopolitan and identified in over 95 % of all samples. The core
311 microbiome represented 0.026 % of bacterial communities and accounted for 2.7 % of all reads. It was
312 composed of 13 OTUs, mainly Proteobacteria and Acidobacteria, notably belonging to the Rhizobiales
313 and Acidobacteria subgroup 6. The Alphaproteobacteria, especially the Bradyrhizobiaceae family was
314 most abundant in acidic soils and decreased with increasing pH. The Betaproteobacteria, especially
315 the Comamonadaceae family, decreased with decreasing pH. The Acidobacteria were highest in the
316 acidoneutral soils. Although relative abundances differed by pH range, these 13 OTUs were present in
317 all categories. Figure S4 illustrates the taxonomic classes of the core microbiome and distribution
318 between each pH category.

319 **Discussion**

320 The coverage of this study allowed the identification of a core microbiome, present in over 95 % of
321 the samples analyzed and composed of only 13 OTUs. The most abundant of these taxa belonged to
322 the Bradyrhizobiaceae family. This is one of the most common families worldwide, as identified by
323 Delgado-Baquerizo et al. (2018). However, whether this is the same taxa remains unclear and
324 highlights the need for global studies incorporating extreme environments. The identification of a
325 core Arctic soil microbiome is novel and this low number of cosmopolitan OTUs illustrates the potential
326 for endemism. For the rest of the bacterial community, spatial and edaphic factors influenced
327 distribution across the region.

328 ***Arctic bacterial community assembly***

329 The distance-decay curve [Fig.2A,2B] illustrated the influence of geography on bacterial
330 communities. The spatial autocorrelation is a proxy for dispersal limitation, which was restricted to
331 approximately 20 m [Fig.2B]. However, the lack of small-scale data is problematic in estimating the real
332 autocorrelation range. The variation partitioning analysis quantified the importance of both,

333 selection (deterministic) and dispersal (stochastic) on the bacterial community structure.
334 Environmental variables explained 30 % of the total variation (selection), of which 15 % was spatially
335 structured, corresponding to the induced spatial dependence, as described in Borcard et al. (2018).
336 Spatial components (trend + MEMs) alone explained 9 % of the variation, illustrating the spatial
337 autocorrelation (Borcard et al., 2018), or dispersal.

338 Here, we focused on environmental variables previously identified as influencing bacterial
339 communities elsewhere and included pH, conductivity, moisture and organic carbon. However, 63 %
340 of the variation remained unexplained by either environmental or spatial factors in the variation
341 partitioning analysis. This unexplained variance could be the result of unmeasured biotic or abiotic
342 factors, directly or indirectly influencing bacterial communities. Examples of unmeasured edaphic
343 properties could include total nitrogen, zinc, calcium oxide or nickel, all shown to have an influence
344 on bacterial community structure in the Arctic (Wojcik et al., 2018). Other ecosystem properties such
345 as ice presence, active layer depth or soil texture are likely to have some influence. Climatic and
346 topographic variables such as temperature, precipitation (Yergeau and Kowalchuk, 2008; Castro et al.,
347 2010; Nielsen and Ball, 2015**)** and altitude (Shen et al., 2013; Xu et al., 2014) are also known to
348 influence bacterial assemblage and were not included in this study. Vegetation cover is a known driver
349 of bacterial community assemblage in the Arctic region (as reviewed by Malard and Pearce (2018))
350 which was not included in this analysis. Biotic interactions may also be a major part of the unexplained
351 variance observed and could include competition and predation within bacterial communities or with
352 other members of the soil biota (Wardle, 2006; Singh et al., 2009). Other processes influencing
353 bacterial communities include diversification and drift, both likely to account for some of the
354 unexplained variance (Martiny et al., 2006).

355 ***The importance of pH***

356 We evaluated the impact of environmental variables known to influence global soil bacterial
357 communities. Overall, pH explained the most variation (18.7 %) in bacterial community structure and

358 diversity while TOC, moisture and conductivity explained between 2 and 5 % of the variance
359 respectively. The identification of pH as a key factor influencing Arctic soil bacterial community
360 composition is in line with previous Arctic studies over both, small and large scales (Männistö et al.,
361 2006; Chu et al., 2010; Ganzert et al., 2014; Siciliano et al., 2014; Schostag et al., 2015). However,
362 this study is the first to quantify the magnitude of influence, highlighting the large variation in
363 bacterial community remaining unexplained. While pH has been identified globally as a major factor
364 influencing microbial diversity and community structure (Fierer and Jackson, 2006; Lauber et al.,
365 2009; Tedersoo et al., 2014; Delgado-Baquerizo et al., 2018), the underlying processes and
366 mechanisms by which it does remain unclear. Studies have demonstrated that the soil pH is correlated
367 with other elements of the geochemistry and has a strong impact on nutrient and water availability as
368 well as solubility and adsorption (Gray et al., 2014). For instance, acidic pH increases aluminium, hydrogen
369 and manganese solubility, retarding plant root growth due to high toxicity (Clark and Baligar, 2000;
370 Singh et al., 2017). Acidic soils also have nutrient deficiencies such as calcium magnesium and
371 potassium but also decreased phosphorus and molybdenum solubilities (Baligar and Clark, 2000; Gray et
372 al., 2014). Alkaline soils are generally the result of low precipitation and high evapotranspiration,
373 leading to low water availability and in common with acidic soils, nutrient deficiencies are found with
374 decreased phosphorus, iron, copper or zinc for instance (Baligar and Clark, 2000). In similar ways,
375 acidic and alkaline soils are generally considered harsh environments requiring a wide range of
376 adaptations from microorganisms while acidoneutral soils are considered the optimum environment
377 for microbial life (Fierer and Jackson, 2006; Rousk et al., 2010); such differences in soil composition
378 are likely driving the observed differences in microbial community composition by pH range.
379
380 ***Distribution of generalist and specialist taxa***

381 Microbial communities are assembled by deterministic (selection) and stochastic (dispersal)
382 processes. It has been hypothesized that communities primarily structured by deterministic processes
383 will host more specialist taxa, highly adapted to the ecosystem, while communities influenced by

384 dispersal will harbour primarily generalist taxa, more resilient to change (Pandit et al., 2009; Graham
385 and Stegen, 2017; Sriswasdi et al., 2017). While specialist taxa are restricted to certain niches, they
386 can be locally abundant; shared taxa, or generalists, are distributed across many niches (Barberán et
387 al., 2012). In most cases specialist will be more abundant because generalists rapidly become
388 specialists to adapt to their ecosystems, despite generalists having evolutionary advantages (Sriswasdi
389 et al., 2017).

390 In this study, bacterial communities were specialist-dominated in acidic soils, generalist-dominated in
391 acidoneutral soils and a mixed community in alkaline samples. The higher abundance of specialists in
392 acidic soils (considered the harshest systems) could illustrate the need for environmental adaptations
393 to survive in these ecosystems and suggests that deterministic processes likely structure microbial
394 communities. Geographically, the first cluster identified by the Bray-Curtis dissimilarity matrix [Fig. 4B]
395 grouped all acidic samples from northern Norway, western Greenland and Alaska, illustrating the
396 similarities of their bacterial communities despite the large distances separating these locations,
397 further suggesting the strong influence of selection over dispersal. The dominance of generalists in
398 acidoneutral soils illustrates the lower environmental pressure to have specific survival adaptations
399 and infers the dominance of stochastic processes in community structuring. Eastern Greenland,
400 Svalbard and Iceland were grouped together by the Bray-Curtis dissimilarity matrix. Geographically,
401 these locations connected through the Greenland sea, a possible route of dispersal through
402 aerosolization and long-distance aeolian transport (Šantl-Temkiv et al., 2018). Finally, alkaline soils
403 hosted a mixed-community of specialists and generalists, suggesting the combination of both,
404 selection and dispersal, in community structuring. The abundance of generalist taxa in alkaline soils,
405 generally considered a harsh system in many ways similar to acidic soils, highlights the adaptability of
406 generalists to a wide range of environmental conditions. Interestingly, the Bray-Curtis dissimilarity
407 matrix clustered the Canadian (alkaline) and Russian (acidoneutral) samples together. The Russian
408 samples were on the higher end of the acidoneutral pH scale and the grouping of these samples
409 illustrates the blurred boundary between acidoneutral and alkaline pH categories. While

410 environmental selection may be driven by many variables, dispersal likely occurs via long-distance
411 aeolian dispersal (Šantl-Temkiv et al., 2018).

412 ***The importance of indicator taxa***

413 The indicator species analysis determined OTU-pH range associations identifying different classes of
414 Acidobacteria and Verrucomicrobia, primarily, characteristic of each pH range. The identification of
415 abundant indicator species with strong habitat associations opens the possibility of predicting the
416 presence and abundance of these taxa across the Arctic region. This is especially important for taxa
417 such as Ca. Methylacidiphilum, a known methanotroph unlike others as it belongs to the
418 Verrucomicrobia phylum instead of the Proteobacteria (Dunfield et al., 2007; Khadem et al., 2010).
419 While natural CH_4 emissions come primarily from wetlands, the identification of these phylotypes in
420 acidic soils suggests wetland may not be the only major source of northern methane emissions.
421 Understanding the distribution and abundance of such indicator taxa, combined with field gas
422 measurements may allow large-scale estimates of climate-active gases release rates in the region,
423 from soils and wetlands (Wartiainen et al., 2006; Jørgensen et al., 2015).

424 ***Concluding remarks***

425 This study investigated patterns in soil bacterial diversity across the Arctic and demonstrated that
426 bacterial communities differed across the region. However, the dispersal range was limited to
427 approximately 20 m and the variation in communities across the Arctic was primarily driven by
428 environmental rather than spatial factors. The key edaphic property influencing communities we
429 identified was pH, in accordance with that seen elsewhere. However, 63 % of the variation remained
430 unexplained, calling for more in-depth investigations of the drivers of bacterial community structure
431 in the region. We identified unique and indicator taxa closely associated to the soil pH. Finally, a core
432 microbiome composed of only 13 ubiquitously distributed taxa was identified, highlighting the
433 potential for endemism in the region. While this study brings a deeper understanding of Arctic
434 bacterial community assemblages, this is also a baseline for future functional studies in the region,

which will be critical to forecast the ecological consequences of environmental change.

Acknowledgements

This work was supported by a grant from the European Commission's Marie Sklowdowska Curie Actions program under project number 675546. The authors also thank Dr Paul Mann, Dr Charles Greer, Dr Svetlana Evgrafova, Dr Bill Holben, Sam Pannoni, 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).

Conflict of interest

The authors report no conflict of interests.

References

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M. et al. (2018) Structure and function of the global topsoil microbiome. *Nature*: 1.

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal* **6**: 343.

Bardgett, R.D., Freeman, C., and Ostle, N.J. (2008) Microbial contributions to climate change through carbon cycle feedbacks. *The ISME journal* **2**: 805.

Borcard, D., Gillet, F., and Legendre, P. (2018) *Numerical ecology with R*: Springer.

Bradford, M.A., Wieder, W.R., Bonan, G.B., Fierer, N., Raymond, P.A., and Crowther, T.W. (2016) Managing uncertainty in soil carbon feedbacks to climate change. *Nature Climate Change* **6**: 751.

Cáceres, M.D., and Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566-3574.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2009) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266-267.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J. et al. (2010a) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*: 201000080.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**: 335.

Castro, H.F., Classen, A.T., Austin, E.E., Norby, R.J., and Schadt, C.W. (2010) Soil microbial community responses to multiple experimental climate change drivers. *Appl Environ Microbiol* **76**: 999-1007.

Chambers, J. (2008) *Software for data analysis: programming with R*: Springer Science & Business Media.

Chu, H., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R., and Grogan, P. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology* **12**: 2998-3006.

Clark, R.B., and Baligar, V.C. (2000) Acidic and alkaline soil constraints on plant mineral nutrition. *Plant-Environment Interactions, 2nd Ed; Wilkinson, RE, Ed*: 133-177.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D. et al. (2018) A global atlas of the dominant bacteria found in soil. *Science* **359**: 320-325.
Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**: 927-930.
Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G. et al. (2017) Adespatial: multivariate multiscale spatial analysis. R package version 0.0-9. In.
Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S. et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**: 879.
Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
Fierer, N. (2008) Microbial biogeography: patterns in microbial diversity across space and time. In *Accessing uncultivated microorganisms*: American Society of Microbiology, pp. 95-115.
Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 626-631.
Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061-1063.
Franklin, R.B., and Mills, A.L. (2007) Statistical analysis of spatial structure in microbial communities. In *The Spatial Distribution of Microbes in the Environment*: Springer, pp. 31-60.
Ganzert, L., Bajerski, F., and Wagner, D. (2014) Bacterial community composition and diversity of five different permafrost-affected soils of Northeast Greenland. *FEMS microbiology ecology* **89**: 426-441.
Graham, D.E., Wallenstein, M.D., Vishnivetskaya, T.A., Waldrop, M.P., Phelps, T.J., Pfiffner, S.M. et al. (2012) Microbes in thawing permafrost: the unknown variable in the climate change equation. *The ISME journal* **6**: 709.
Graham, E., and Stegen, J. (2017) Dispersal-based microbial community assembly decreases biogeochemical function. *Processes* **5**: 65.
Gray, N., McCann, C., Christgen, B., Ahammad, S., Roberts, J., and Graham, D. (2014) Soil geochemistry confines microbial abundances across an arctic landscape; implications for net carbon exchange with the atmosphere. *Biogeochemistry* **120**: 307-317.
Henschel, A., Anwar, M.Z., and Manohar, V. (2015) Comprehensive meta-analysis of ontology annotated 16S rRNA profiles identifies beta diversity clusters of environmental bacterial communities. *PLoS computational biology* **11**: e1004468.
Jørgensen, C.J., Johansen, K.M.L., Westergaard-Nielsen, A., and Elberling, B. (2015) Net regional methane sink in High Arctic soils of northeast Greenland. *Nature Geoscience* **8**: 20.
Khadem, A.F., Pol, A., Jetten, M.S., and den Camp, H.J.O. (2010) Nitrogen fixation by the verrucomicrobial methanotroph 'Methylacidiphilum fumariolicum'SolV. *Microbiology* **156**: 1052-1059.
Kleinteich, J., Hildebrand, F., Bahram, M., Voigt, A.Y., Wood, S.A., Jungblut, A.D. et al. (2017) Pole-to-pole connections: Similarities between Arctic and Antarctic microbiomes and their vulnerability to environmental change. *Frontiers in Ecology and Evolution* **5**: 137.
Koven, C.D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D. et al. (2011) Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences* **108**: 14769-14774.
Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*: AEM. 01043-01013.
Lanzén, A., Jørgensen, S.L., Huson, D.H., Gorfer, M., Grindhaug, S.H., Jonassen, I. et al. (2012) CREST–classification resources for environmental sequence tags. *PloS one* **7**: e49334.
Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and environmental microbiology* **75**: 5111-5120.
Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **44**: W242-W245.

Ma, W.K., Schautz, A., Fishback, L.-A.E., Bedard-Haughn, A., Farrell, R.E., and Siciliano, S.D. (2007) Assessing the potential of ammonia oxidizing bacteria to produce nitrous oxide in soils of a high arctic lowland ecosystem on Devon Island, Canada. *Soil Biology and Biochemistry* **39**: 2001-2013.

Mackelprang, R., Waldrop, M.P., DeAngelis, K.M., David, M.M., Chavarria, K.L., Blazewicz, S.J. et al. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**: 368-371.

Magoč, T., and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957-2963.

Malard, L.A., and Pearce, D.A. (2018) Microbial diversity and biogeography in Arctic soils. *Environmental microbiology reports*.

Männistö, M.K., Tiirola, M., and Häggblom, M.M. (2006) Bacterial communities in Arctic fjelds of Finnish Lapland are stable but highly pH-dependent. *FEMS Microbiology Ecology* **59**: 452-465.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**: 10-12.

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L. et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* **4**: 102-112.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**: e61217.

Mittelbach, G.G., and Schemske, D.W. (2015) Ecological and evolutionary perspectives on community assembly. *Trends in ecology & evolution* **30**: 241-247.

Nazaries, L., Murrell, J.C., Millard, P., Baggs, L., and Singh, B.K. (2013) Methane, microbes and models: fundamental understanding of the soil methane cycle for future predictions. *Environmental microbiology* **15**: 2395-2417.

Nielsen, U.N., and Ball, B.A. (2015) Impacts of altered precipitation regimes on soil communities and biogeochemistry in arid and semi-arid ecosystems. *Global Change Biology* **21**: 1407-1421.

O'Malley, M.A. (2007) The nineteenth century roots of'everything is everywhere'. *Nature Reviews Microbiology* **5**: 647-651.

Pandit, S.N., Kolasa, J., and Cottenie, K. (2009) Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* **90**: 2253-2262.

Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013) Differential abundance analysis for microbial marker-gene surveys. *Nature methods* **10**: 1200.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **5**: e9490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* **35**: 7188-7196.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G. et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME journal* **4**: 1340-1351.

Šantl-Temkiv, T., Gosewinkel, U., Starnawski, P., Lever, M., and Finster, K. (2018) Aeolian dispersal of bacteria in southwest Greenland: their sources, abundance, diversity and physiological states. *FEMS microbiology ecology* **94**: fiy031.

Schostag, M., Stibal, M., Jacobsen, C.S., Bælum, J., Taş, N., Elberling, B. et al. (2015) Distinct summer and winter bacterial communities in the active layer of Svalbard permafrost revealed by DNA-and RNA-based analyses. *Frontiers in microbiology* **6**: 399.

Schuur, E., McGuire, A.D., Schädel, C., Grosse, G., Harden, J., Hayes, D. et al. (2015) Climate change and the permafrost carbon feedback. *Nature* **520**: 171-179.

Shen, C., Xiong, J., Zhang, H., Feng, Y., Lin, X., Li, X. et al. (2013) Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. *Soil Biology and Biochemistry* **57**: 204-211.

Siciliano, S.D., Palmer, A.S., Winsley, T., Lamb, E., Bissett, A., Brown, M.V. et al. (2014) Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biology and Biochemistry* **78**: 10-20.

Singh, B.K., Dawson, L.A., Macdonald, C.A., and Buckland, S.M. (2009) Impact of biotic and abiotic interaction on soil microbial communities and functions: A field study. *Applied Soil Ecology* **41**: 239-248.

Singh, S., Tripathi, D.K., Singh, S., Sharma, S., Dubey, N.K., Chauhan, D.K., and Vaculík, M. (2017) Toxicity of aluminium on various levels of plant cells and organism: a review. *Environmental and Experimental Botany* **137**: 177-193.

Sriswasdi, S., Yang, C.-c., and Iwasaki, W. (2017) Generalist species drive microbial dispersion and evolution. *Nature communications* **8**: 1162.

Team, R.C. (2013) R: A language and environment for statistical computing.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R. et al. (2014) Global diversity and geography of soil fungi. *science* **346**: 1256688.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J. et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**.

Wardle, D.A. (2006) The influence of biotic interactions on soil biodiversity. *Ecology letters* **9**: 870-886.

Wartiainen, I., Hestnes, A.G., McDonald, I.R., and Svenning, M.M. (2006) Methylobacter tundripaludum sp. nov., a methane-oxidizing bacterium from Arctic wetland soil on the Svalbard islands, Norway (78 N). *International Journal of Systematic and Evolutionary Microbiology* **56**: 109-113.

Wojcik, R., Donhauser, J., Frey, B., Holm, S., Holland, A., Anesio, A.M. et al. (2018) Linkages between geochemistry and microbiology in a proglacial terrain in the High Arctic. *Annals of Glaciology* **59**: 95-110.

Xu, M., Li, X., Cai, X., Gai, J., Li, X., Christie, P., and Zhang, J. (2014) Soil microbial community structure and activity along a montane elevational gradient on the Tibetan Plateau. *European journal of soil biology* **64**: 6-14.

Yergeau, E., and Kowalchuk, G.A. (2008) Responses of Antarctic soil microbial communities and associated functions to temperature and freeze–thaw cycle frequency. *Environmental microbiology* **10**: 2223-2235.

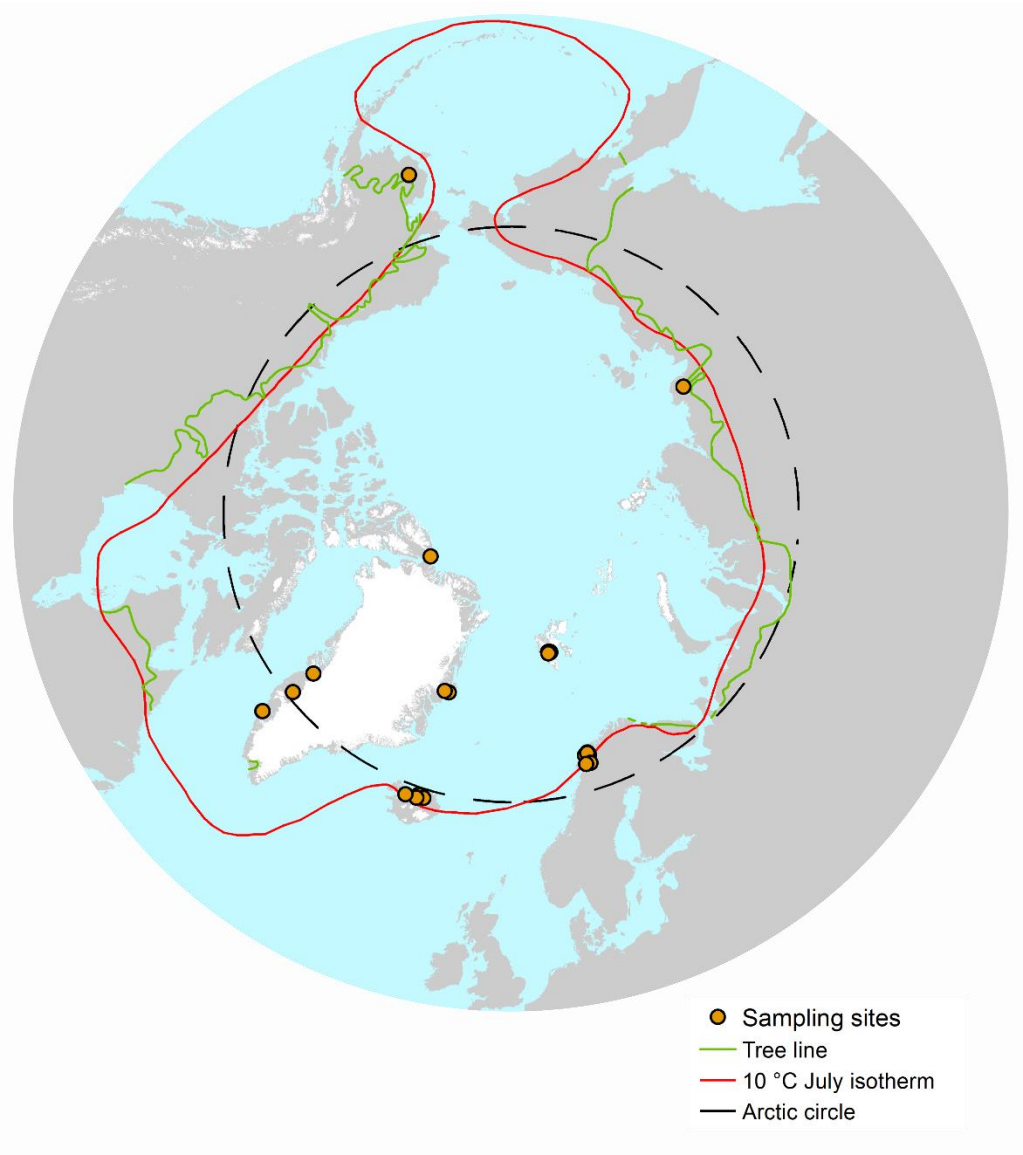
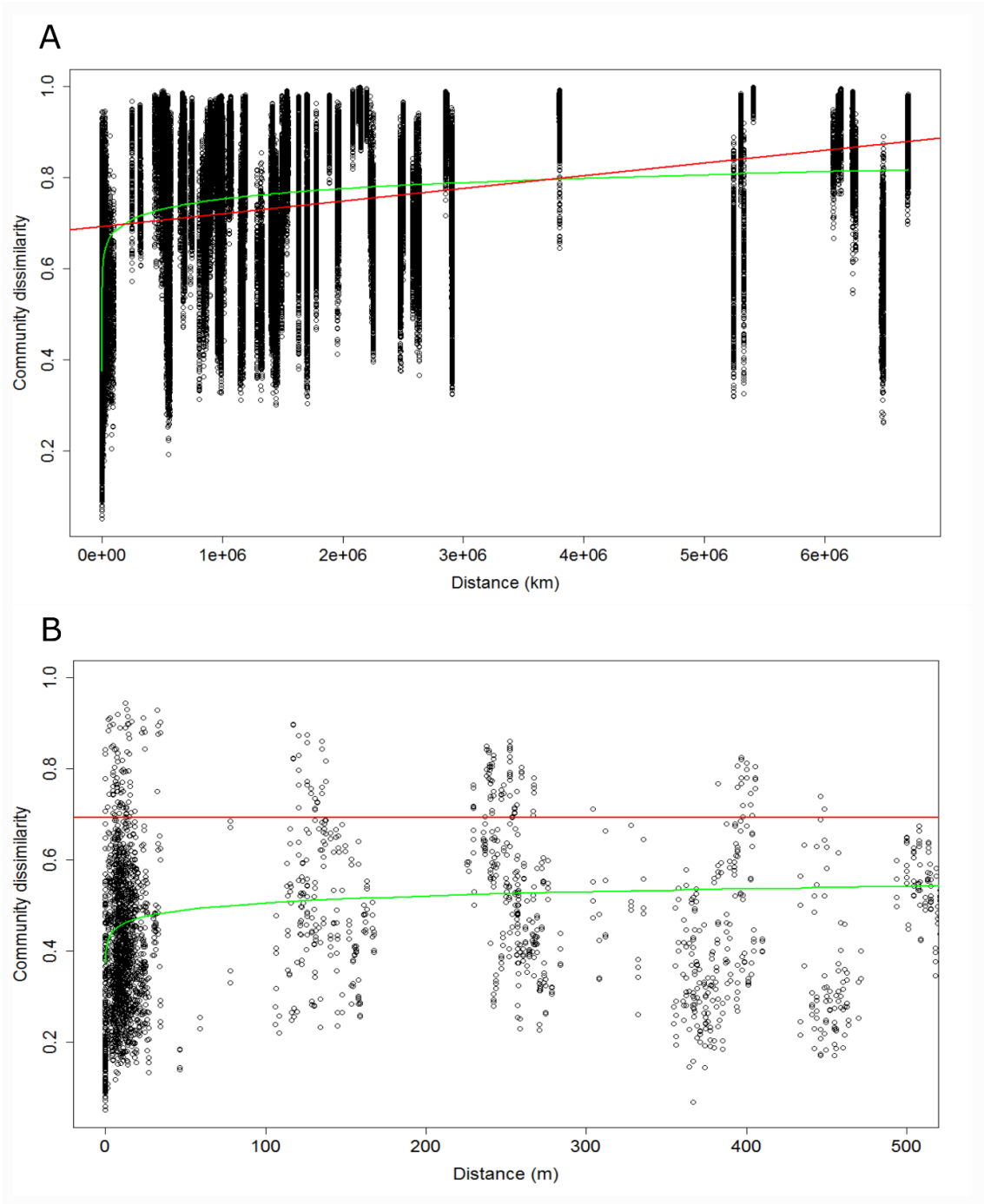


Figure 1: Map of sampling sites within the Arctic Region, as characterised by the Arctic circle (black line), the 10 °C isotherm (red) and the tree line boundary (green). 200 individual soil samples at 43 different sites were collected for this study, all within at least one of the definition of the Arctic region.



611

612 *Figure 2: Distance-decay curves with the linear regression (red), the best-fitted power model (green).*

613 *(A) Community dissimilarity of the total bacterial community by geographic distance in km. (B) Zoom*

614 *into Figure 2A from 0 to 500 m to assess the spatial autocorrelation distance.*

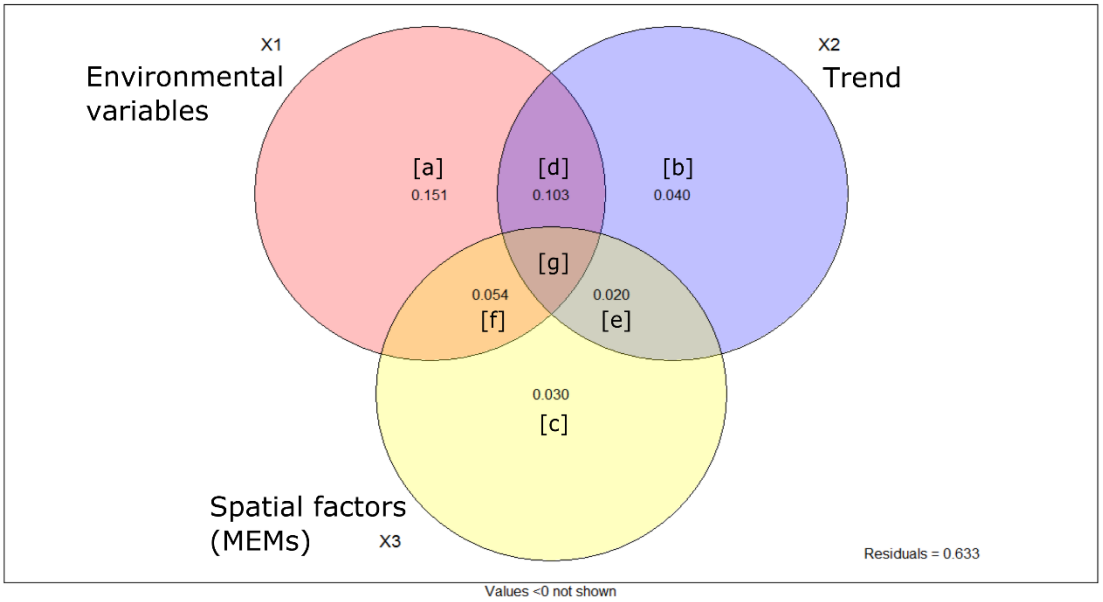


Figure 3: Venn diagram illustrating the results of the variation partitioning analysis and indicating the amount of variation in bacterial community explained by environmental variables and spatial factors including the linear trend. Detailed results are available in Table S1.

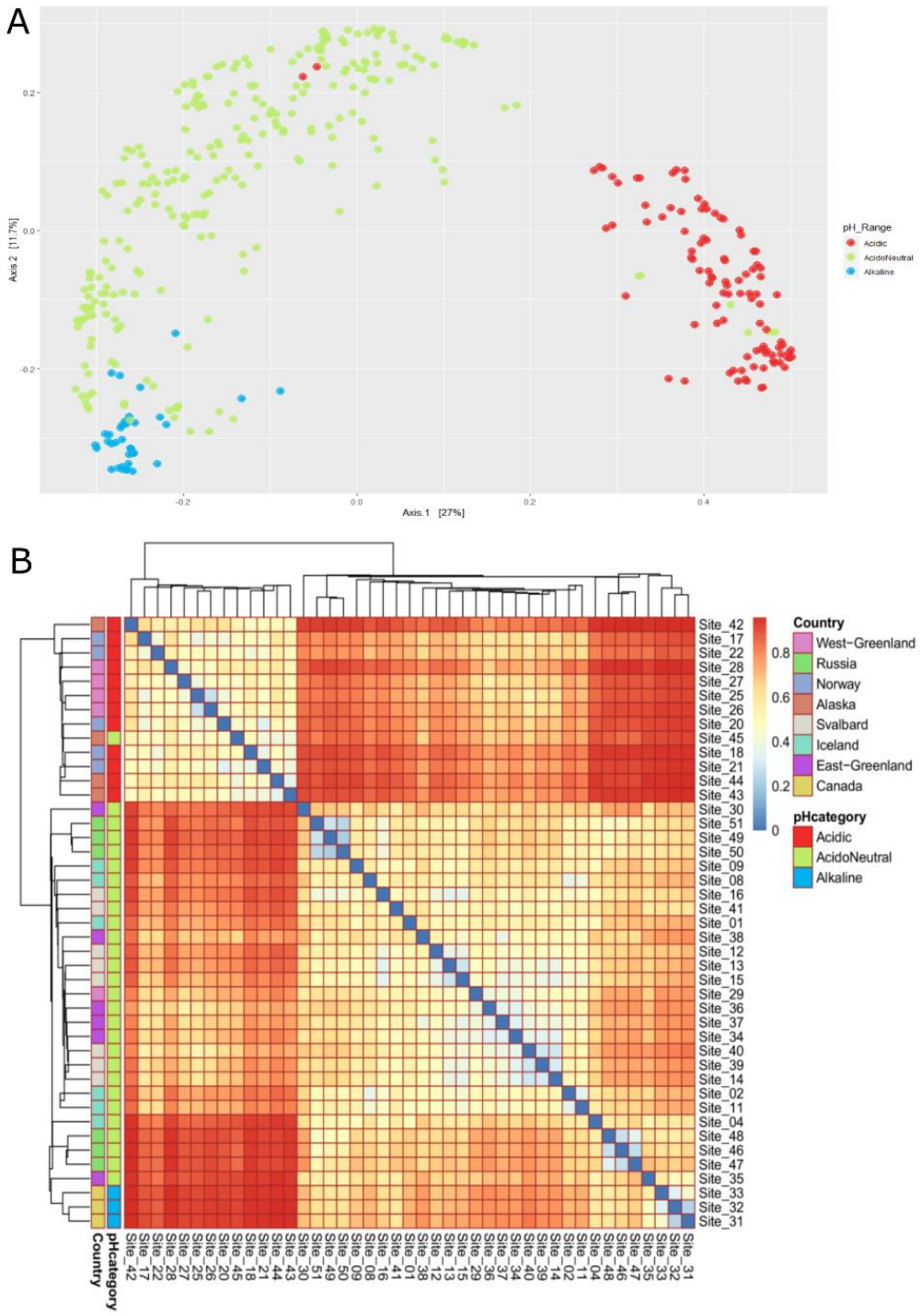


Figure 4: (A) PCoA of the Bray-Curtis dissimilarity of bacterial communities across the Arctic region by pH range. (B) Bray-Curtis dissimilarity heatmap and dendrogram illustrating the clustering of samples into three pH categories.

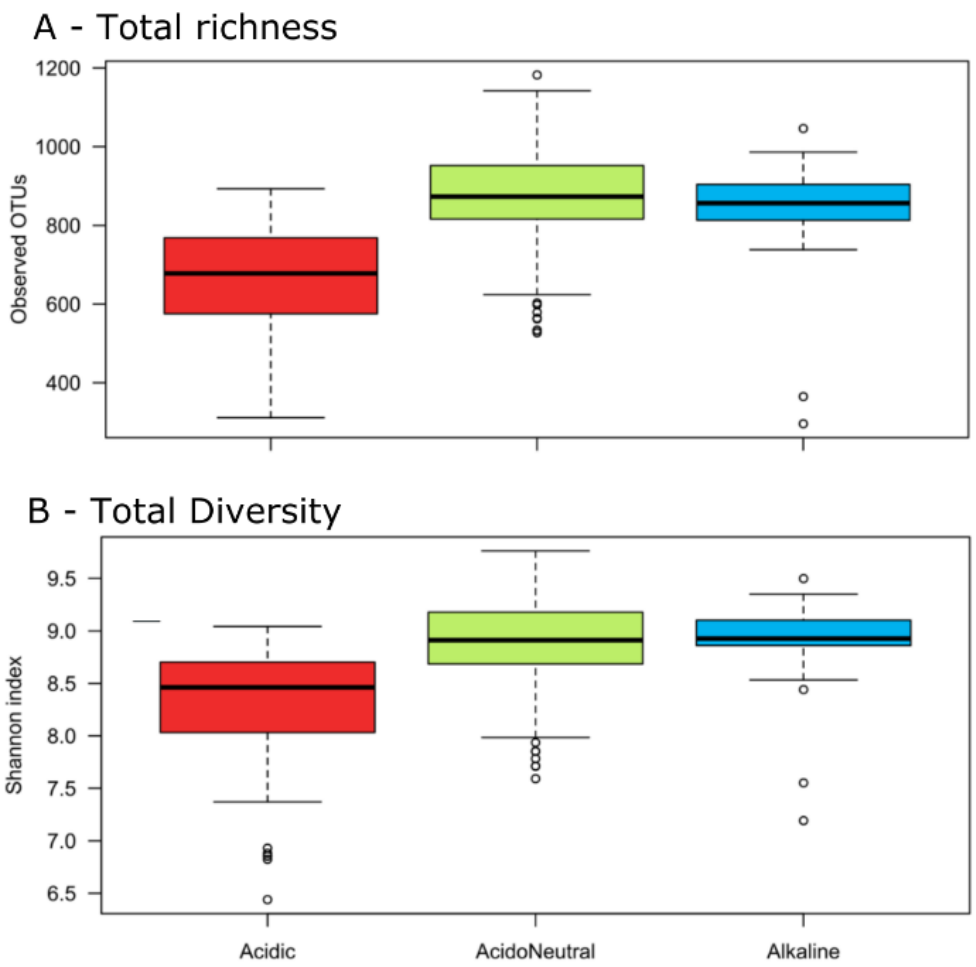


Figure 5: Alpha diversity differences in the total communities by pH category. (A) Richness by observed OTUs. (B) Shannon index.

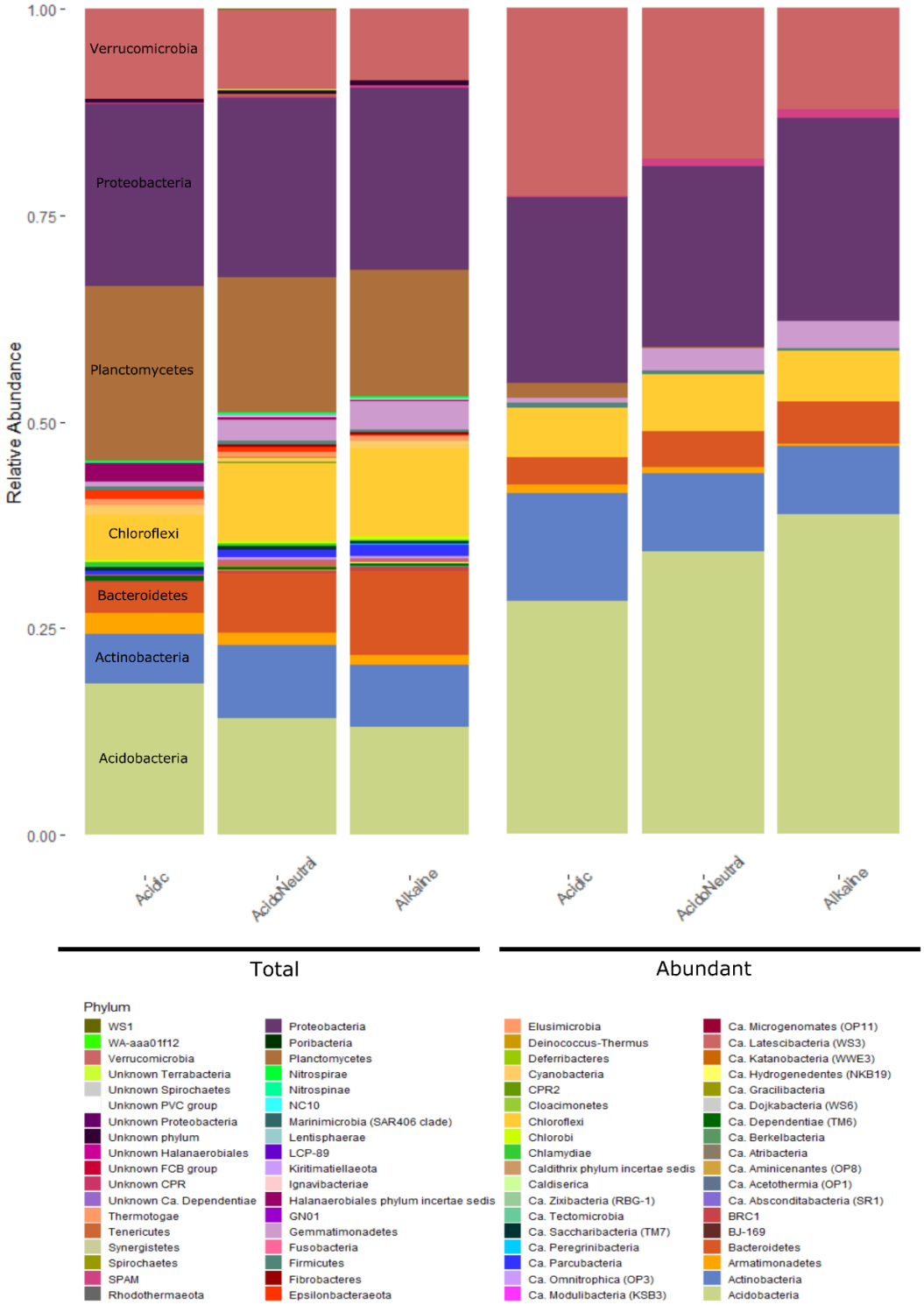


Figure 6: Relative abundance of the total and abundant communities by pH range classified at the phylum level.

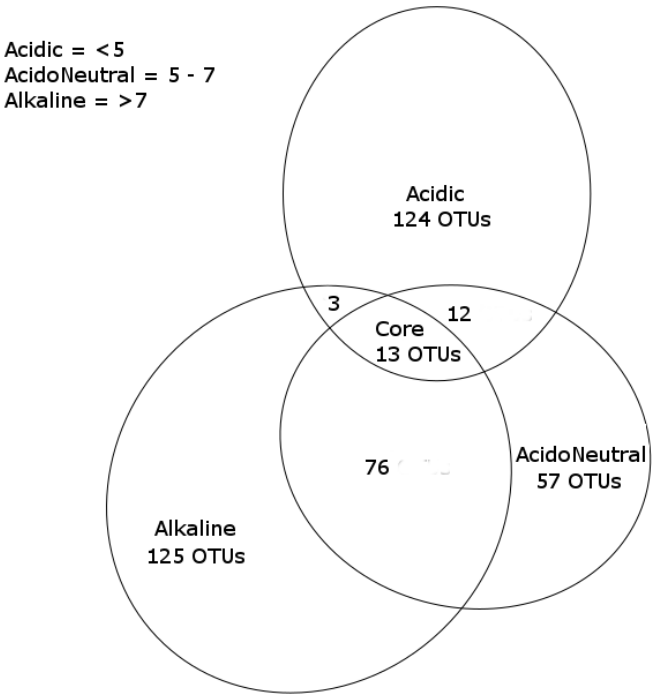


Figure 7: Venn diagram illustrating generalist (shared) and specialist (unique) OTUs by pH category.

Table 1: The relative importance of environmental factors on bacterial communities, calculated by adonis.

	Total	
Variable	**R^2**	**Pr(>F)**
pH	0.187	0.0001***
Conductivity	0.048	0.0001***
Moisture	0.027	0.0001***
TOC	0.021	0.0001***
Residuals	0.717	N/A