

Quebec, 2 December 2015

Isabelle Laurion,
Associate editor - Biogeosciences,
European Geophysical Union

Dear Dr. Laurion,

RE: Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes by J. Comte, C. Lovejoy, S. Crevecoeur, and W.F. Vincent

Thank you very much for your consideration and your detailed comments. Please find attached our revised materials.

All suggestions have now been incorporated in this revision, and our detailed responses to comments are given below. Line numbers in the responses refer to the final copy.

We thank the Editor for her insightful contribution and interest in our work. We reconfirm that this material has not been published or submitted for publication elsewhere, the content and authorship of the submitted manuscript has been approved by the co-authors and there is no conflict of interest.

Sincerely,

Jérôme Comte, PhD
Département de biologie - Université Laval
1030, avenue de la Médecine
Québec (Québec) Canada
G1V 0A6

Page 3

Lines 45-46: the argument behind mentioning this is unclear: is it because this stock of OC once dissolved in lake water can drive the bacterial communities as evidenced by your study?

The goal of this sentence was just to convey to the reader to the large amount of C stored in permafrost landscape and to highlight its relevance at the global scale.

Lines 52-54: this part of sentence is only supported by Walter et al. 2008, not by Laurion et al. 2010 neither by Negandhi et al 2013. And in fact I don't think it's right to say they are net sources of GHG to the atmosphere DUE to the mobilization of ancient C. There is still a lot of work to be done to quantify to which degree the old stock is fueling microbes. And see Bouchard et al. in the special issue related to this question

We agree with the Editor that the direct link between mobilization of ancient C stored in permafrost and emission of green gas is still under discussion. Therefore, we have altered the text that now reads: “Although some are carbon sinks (Walter Anthony et al., 2014),

others are net sources of carbon dioxide (CO₂) and methane (CH₄) to the atmosphere (Walter et al., 2008)". We have also removed the references mentioned by the Editor.

Page 4

Line 75: reverse so that we can understand you mean bacterial distribution, otherwise it becomes the distribution of lakes and ponds

We have altered the text that now reads: "[...] conditions are likely to influence the distribution of bacterial communities of thaw ponds and lakes."

Line 94: why autotrophic and not phototrophic?

We agree with the Editor that autotrophic was too broad and our data mainly targeted the phototrophs (fluorescence based on their pigment content). We have changed autotrophic to phototrophic throughout the manuscript.

Page 5

Line 114: capital letter not needed

Now corrected.

Line 115: strictly speaking, water for DOC is filtered THROUGH a filter, and water for Chla is filtrered ONTO a filter... but I understand you were trying to condense...

We have changed the text accordingly

Line 126: I don't think this informatiion is necessary, rarely the lab is given, unless for very specific analyses

We have removed this information.

Line 128: strictly speaking this should be absorbance, that is used in the SUVA calculation (not absorption coefficient)

We have changed this accordingly

Page 6

Line 133: remove parenthesis in front of author

Now corrected.

Line 146: but I beleive this is done on UNstained samples right? (no SYBR Green)

The sentence now reads: "[...] whereas phototrophic picoeukaryotes and picocyanobacteria were discriminated from unstained samples on the basis of their red autofluorescence...".

Lines 157 and 443-450: I suggest that you make this choice more explicit, which means that you excluded bacterial communities associated to particles (collected on the 3um filters), and this can be a large fraction in thermokarst lakes... You might want to briefly acknowledge this in the discussion as well.

We added the following sentences:

In the Methods : "For this study, the bacterial community composition of the free-living fraction (< 3 μm) was examined."

In the Discussion: “The patterns described here are for the free-living fraction of bacterial assemblages, which raises the question of whether such patterns remain for the attached fraction of the communities. The latter may represent a substantial part of the total communities given that these waterbodies can contain a large content of suspended solids. Previous studies comparing the compositional patterns in bacterial communities between the free-living and attached fractions showed that these two distinct life-style have a similar community composition (Crevecoeur et al. 2015), indicating that the patterns described here may reflect patterns for the entire community..”

Page 9

Line 249: capital letter not necessary

Now modified.

Lines 250-252: these are also abundances (not only bacterial)

We have specified that these variables represented abundance data.

Page 15

Here and below down to the end of conclusion: I understand that when you talk in general (intro), you want to include both lakes and ponds, but in your dataset, can't you decide if your water bodies are lakes or ponds according to your classification (freezing to bottom?, connected to hydrographic network? size?) instead of carrying this unprecise/large qualification?

We used both terms in addition to “thermokarst systems” given that the terminology used to determine these water bodies is still under discussion. In addition, our study may include both ponds and lakes as we sampled a large number of heterogeneous aquatic systems. Unless the Editor advises us otherwise, we would prefer not change the terminology used in the manuscript.

Page 16

I am searching for a description of this specificity in DOC (or you mean in DOM i.e. DOC and SUVA?) at SAS, but cannot find it in the results.

We did observe significant difference in SUVA and DOC concentration among the different valleys and where ponds from SAS valley have higher values on average. These data were not shown in this manuscript but previous studies on these systems showed that DOC was a clear determinant for community composition of SAS ponds and lakes (Comte et al. 2015, Crevecoeur et al. 2015). The argument we make here is that the uniqueness related to SAS ponds and lakes may be related to the palsa origin (organic versus mineral), which may likely be associated with differences in the DOM molecular composition, which we did not characterize in this study.

Page 17

Lines 471-477: I think you are not explicit enough on why you make a link to methanotrophs; I suggest this part makes its own sentence with the link to specificity of SAS DOM and CH₄ (more CH₄ brings more CH₄ consumers I suppose?).

We have altered the text following reviewer’s suggestion, which now reads: “Consistent with this pattern, elevated concentrations and high rates of CO₂ and CH₄ emission to the

atmosphere have been reported among the southern sites within the most degraded area of permafrost (Laurion et al., 2010; Deshpande et al., 2015). This may in turn explain the significantly higher bacterial richness and diversity observed in SAS thaw ponds communities and why OTUs assigned to methanotrophic bacteria such as *Methylobacter* and *Methylothera* were amongst the most abundant detected in this valley (Fig. 2)".

Lines 501-504: is this bringing lower diversity in general/elsewhere?

The point here was that the reference lakes were more similar to each others than with the thaw lakes, possibly because they have been subjected to longer evolutionary processes than the thaw lakes that are "younger" systems.

Line 504: is this equivalent in saying lower diversity, i.e. more similar? I am confused now cause you said above rock ponds had more similar communities (than thaw ponds right?)

We agree with the Editor that this was confusing. We have removed the last sentence "In contrast, the high environmental heterogeneity in the permafrost landscape is consistent with the higher degree of community relatedness observed among thaw ponds".

Page 18

Line 506-510: the 2 "than" makes the comparison confusing

The sentence now reads: "When controlling for the two outliers mentioned above (NAS-A and NAS-B), the northern communities (BGR, NAS) had a greater phylogenetic distance among co-occurring taxa than expected by chance (lower NRIs) in comparison to communities from the thaw ponds located in valleys from sporadic permafrost (KWK, SAS)".

Lines 510-513: please specify what constrains are you referring to for SAS?

The text now reads: "This suggests that taxa from SAS valley (and to a lesser extent KWK), tend to be more ecologically similar to each other than those from northern valleys, reflecting strong environmental filtering by variables such as DOC concentration, as previously documented in this valley (Comte et al., 2015)".

Page 20

Line 572: remove parentheses?

Now corrected.

Line 581: highly?

Now corrected.

Lines 578-582-628: what do you mean? less likely driven by this in northern site?

Yes, Fig.2A shows that the relationship between OTUs connectivity and the degree of specificity of the OTUs is strong for those retrieved in ponds located within the southern valleys.

Line 586: is this an expression used in other context or by Montoya? If so, please define further what you mean, otherwise useless

This is a commonly used term in network theory and means that the networks are dominated by a few number of highly connected OTUs or species.

Lines 601-602: it would be interesting to discuss why N-fixing bacterium would be important in an environment relatively rich in N

We revised the text that now reads: “This in turn highlights the potentially important ecological role of diazotrophs in these nutrient-rich aquatic systems”.

Page 21

Line 606: you mean among lakes?

Yes, we have changed “sites” to “waterbodies”.

Page 22

Lines 633-634: All first names could be abbreviated (as above) (same for Marie-Josée and Isabelle)

Now corrected.

Line 634: I think there is no need to thank INRS for analyses that were paid.

Now deleted.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

**Co-occurrence patterns in aquatic bacterial communities across changing
permafrost landscapes**

J. Comte ^{1,2}, C. Lovejoy, ^{1,2,3}, S. Crevecoeur ^{1,2}, and W. F. Vincent ¹

¹Centre d'études nordiques (CEN), Takuvik Joint International Laboratory &
Département de biologie, Université Laval, Québec, QC G1V 0A6, Canada

²Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC
G1V 0A6, Canada.

³Québec Océan, Université Laval, Québec, QC G1V 0A6, Canada.

Special issue, Freshwater ecosystems in changing permafrost landscapes

Author for correspondence: Jérôme Comte, email: jerome.comte@takuvik.ulaval.ca

19 **Abstract**

20 Permafrost thaw ponds and lakes are widespread across the northern landscape and may
21 play a central role in global biogeochemical cycles, yet knowledge about their microbial
22 ecology is limited. We sampled a set of thaw ponds and lakes as well as shallow rock-
23 basin lakes that are located in distinct valleys along a North-South permafrost
24 degradation gradient. We applied high-throughput sequencing of the 16S rRNA gene to
25 determine co-occurrence patterns among bacterial taxa (operational taxonomic units,
26 OTUs), and then analyzed these results relative to environmental variables to identify
27 variables controlling bacterial community structure. Network analysis was applied to
28 identify possible ecological linkages among the bacterial taxa and with abiotic and biotic
29 variables. The results showed an overall high level of shared taxa among bacterial
30 communities within each valley, however the bacterial co-occurrence patterns were non-
31 random, with evidence of habitat preferences. There were taxonomic differences in
32 bacterial assemblages among the different valleys that were statistically related to
33 dissolved organic carbon concentration, conductivity and phytoplankton biomass. Co-
34 occurrence networks revealed complex interdependencies within the bacterioplankton
35 communities and showed contrasting linkages to environmental conditions among the
36 main bacterial phyla. The thaw pond networks were composed of a limited number of
37 highly connected taxa. This ‘small world network’ property would render the
38 communities more robust to environmental change but vulnerable to the loss of microbial
39 keystone species. These highly connected nodes (OTUs) in the network were not merely
40 the numerically dominant taxa, whose loss would greatly alter the organization of
41 microbial consortia and ultimately the food web structure and functioning of these
42 aquatic ecosystems.

43

44 1 Introduction

45 Permafrost is widespread in Arctic and boreal regions (Schuur et al., 2008) and is
46 estimated to contain ca. 1700 Pg of organic carbon (McGuire et al., 2009; Tarnocai et al.,
47 2009). Permafrost thawing and erosion is evident by the northward retreat of the
48 permafrost boundary (Thibault and Payette, 2009). In some northern regions this has led
49 to the expansion of permafrost thaw ponds and lakes (thermokarst systems; Grosse et al.,
50 2013), whereas in other regions there has been a contraction and loss of these waterbodies
51 (e.g., Andresen and Lougheed, 2015). These thermokarst systems are part of circumpolar
52 and global biogeochemical cycles (Abnizova et al., 2012; Walter et al., 2007). **Although**
53 **some are carbon sinks (Walter Anthony et al., 2014), others are net sources of carbon**
54 **dioxide (CO₂) and methane (CH₄) to the atmosphere (Walter et al., 2008).**

55 Bacterial communities are among the main drivers of key biogeochemical processes
56 (Ducklow, 2008), and in thermokarst systems are composed of functionally diverse taxa
57 (Crevecoeur et al., 2015; Rossi et al., 2013). In particular, these systems are favorable for
58 bacterial methanotrophs (Crevecoeur et al., 2015) as well as archaeal methanogens
59 (Mondav et al., 2014), and the relative activity of these two groups will affect methane
60 balance and the net emission of greenhouse gases. Identifying factors that shape bacterial
61 communities in these aquatic systems is therefore essential for understanding the
62 functional significance of these permafrost thaw systems in the global carbon budget.

63 Aquatic bacterial communities are thought to be selected by a combination of bottom-
64 up (resource availability) and top-down (viral lysis, grazing) controls. Less studied are
65 bacteria-bacteria interactions (facilitation, competition), which may further contribute to
66 non-random distributions observed among microbial taxa (e.g., Horner-Devine et al.,
67 2007). Examining co-occurrence patterns has the potential to unveil ecological processes
68 that structure bacterial communities. Specifically, patterns of co-occurrence may reveal to
69 what extent groups of microbes share habitat preferences, to what extent there may be
70 ecological linkages among bacterial taxa and with other planktonic organisms, and the
71 extent of phylogenetic closeness of co-occurring bacterial taxa given that closely related
72 taxa may share life strategies and ecological traits.

73 **Across northern landscapes, both regional (e.g., climate and the degradation state of**
74 **permafrost) and local (e.g., nutrients, dissolved organic carbon and oxygen) conditions**

75 are likely to influence the distribution of bacterial communities of thaw ponds and lakes.
76 These thermokarst systems show a high degree of limnological (Deshpande et al., 2015)
77 and bacterial heterogeneity (Crevecoeur et al., 2015), making them suitable models to
78 investigate the co-occurrence patterns among bacterial taxa as well their network
79 relationships within microbial consortia. The main objectives of this study were to
80 characterize the ecological linkages within microbial communities as a response to
81 permafrost thawing. Our hypotheses were that (i) bacterial communities follow co-
82 occurrence patterns along the permafrost degradation gradient, due to distinct habitat
83 preferences among bacteria, and (ii) these habitat preferences relate to differences in the
84 phylogenetic structure of bacterial communities.

85 To test the above hypotheses, we employed high-throughput sequencing of the 16S
86 rRNA gene to determine the composition of bacterial communities in thaw ponds and
87 lakes of Nunavik (Quebec, Canada) along a North-South permafrost degradation
88 gradient. In addition, we sampled rock-basin lakes that were under the same regional
89 climate but whose formation was not related to climate change. We investigated the
90 relationships among bacterial taxa and local environmental conditions by means of
91 network analysis, which has been applied with success elsewhere to evaluate microbial
92 distribution patterns (Barberan et al., 2012; Peura et al., 2015; Steele et al., 2011) and
93 responses to environmental perturbation (Araújo et al., 2011). We then examined the
94 potential linkages between the bacteria and phytoplankton, phototrophic picoplankton
95 and zooplankton biomass in the ponds.

97 **2 Methods**

98 **2.1 Study sites and sampling**

99 Surface water (0.2 m) from 29 thermokarst ponds was collected from 1 to 13 August
100 2012 in two types of permafrost landscapes. Thaw ponds were located in the vicinity of
101 Whapmagoostui-Kuujuarapik (W-K: lat. 55° 15' N, long. 77° 45' W) and Umiujaq (lat.
102 56° 32' N, long. 76° 33' W), within four valleys in the eastern Canadian subarctic,
103 Nunavik along a North-South permafrost degradation gradient as described in Comte et
104 al. (2015): the Sasapimakwananisikw River valley (SAS) and the Kwakwatanikapistikw

105 River valley (KWK), in sporadic, highly degraded permafrost landscapes (< 10%
106 permafrost coverage; see Bhiry et al., 2011 for details); and the Sheldrake River valley
107 (BGR) and Nastapoka River valley (NAS) that are in discontinuous permafrost
108 landscapes (10-50% permafrost coverage). In addition, we sampled 5 rock-basin lakes as
109 ‘reference lakes’ (RBL) in catchments near the W-K village as a fifth ‘valley’; these
110 waters occupy glacially scoured basins, and their origin is not related to permafrost
111 degradation.

112 At each site, temperature, conductivity, dissolved oxygen and pH were measured using
113 a 600R multiparametric probe (YSI, Yellow Springs, OH, USA). Water for dissolved
114 organic carbon (DOC) and **chlorophyll-*a*** (Chl-*a*) was filtered through MilliQ water pre-
115 rinsed 47-mm diameter, 0.22- μ m pore size acetate filters and **onto** GF/F filters,
116 respectively (Whatman, GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire,
117 UK). Water samples for total phosphorus (TP) and total nitrogen (TN) were preserved
118 with H₂SO₄ (0.15% final concentration) until further analyses.

119 Samples for zooplankton were collected using a 35 μ m net, fixed in ethanol (final
120 concentration: 75%, v/v), and stored in cold (4 °C) dark conditions until analysis by
121 inverted microscopy. Microbial abundance samples for flow cytometry (FCM) analysis
122 were further collected and fixed with glutaraldehyde (final concentration: 2%, v/v) and
123 stored frozen at -80 °C until analysis.

124

125 **2.2 Chemical and plankton analyses**

126 DOC concentrations were analyzed on a Shimadzu TOC-5000A carbon analyzer and
127 nutrients were analyzed using standard methods (Stainton et al., 1977). Colored dissolved
128 organic matter (CDOM) was measured by spectrophotometric analysis of **absorbance** at
129 254 nm by water filtered through 0.2 μ m pore-size filters and the dissolved aromatic
130 carbon content was determined using the SUVA₂₅₄ index (Weishaar et al., 2003).

131 Phytoplankton biomass was estimated as chlorophyll *a* concentrations (Chl-*a*), which
132 were determined using high performance liquid chromatography (ProStar HPLC system,
133 Varian, Palo Alto, CA, USA) following the procedures described in **Bonilla** et al. (2005).
134 Zooplankton, specifically copepods, rotifers and cladocerans, were enumerated following
135 the Utermöhl procedure (1958) and inverted microscopy (Zeiss Axiovert, Carl Zeiss

136 Microscopy GmbH, Jena, Germany). Bacteria, picocyanobacteria and phototrophic
137 picoeukaryotes were enumerated using a FACScalibur flow cytometer (BD, Mississauga,
138 ON, Canada), equipped with an argon laser, at the lowest flow rate ($12 \mu\text{l min}^{-1}$), using 1
139 μm yellow green microspheres (Polysciences Inc, Warrington, PA, USA) in suspension
140 as an internal standard. Bead concentration was controlled using Truecount Absolute
141 counting tubes (BD, Mississauga, ON, Canada). Bacteria were stained by adding 20 μl of
142 a 50X SYBR Green I (Life Technologies, Thermo Fisher Scientific, Waltham, MA,
143 USA) to 500 μl of sample for 10 min in the dark. Bacterial cells were then discriminated
144 on the basis of their green fluorescence (FL1) and side scatter signals (SSC) while excited
145 at 488 nm, whereas phototrophic picoeukaryotes and picocyanobacteria were
146 **discriminated from unstained samples on the basis of their red autofluorescence** (FL3)
147 with a threshold in orange (FL2) and SSC. The resulting data were analyzed using the
148 CellQuest Pro software with manual gating.

149

150 **2.3 Bacterial community composition**

151 Bacterial community composition (BCC) was determined by 454-pyrosequencing of
152 the V6-V8 regions of the 16S rRNA gene. In brief, water was sequentially filtered
153 through a 20 μm mesh net to remove larger organisms, a 47-mm diameter, 3 μm pore size
154 polycarbonate filter (Whatman) and a 0.2 μm pore size Sterivex unit (EMD Millipore,
155 Billerica, MA, USA) using a peristaltic pump. The filters were preserved with 1.8 ml of
156 *RNAlater* (Life Technologies) and stored at -80°C until further processing. **For this**
157 **study, the bacterial community composition of the free-living fraction ($< 3 \mu\text{m}$) was**
158 **examined.** DNA was extracted from cells collected onto Sterivex units using the
159 PowerWater Sterivex DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA
160 USA) following the manufacturer instructions. Extracted DNA was amplified in three
161 separate 20 μl PCR reactions using 1 μl of template (3 concentrations: 1, 0.5, and 0.2X)
162 and a Phusion high-fidelity DNA polymerase kit (New England Biolabs, Whitby, ON,
163 USA), and reverse 1406R and forward 969F primers with sample specific tags as in
164 Comeau et al. (2011). Amplicons were purified using a PCR purification kit from Feldan
165 (QC, Canada), quantified spectrophotometrically (Nanodrop, ND-1000, Wilmington, DE,
166 USA) and sequenced using Roche/454 GS FLX Titanium technology at Plateforme

167 d'Analyses Génomiques, Institut de Biologie Intégrative et des Systèmes, Université
168 Laval (Québec, Canada). The raw reads have been deposited in the NCBI database under
169 the accession number SRP044372.

170 All sequence data processing was within the QIIME v1.8.0 pipeline (Caporaso et al.,
171 2010b). Reads were first pre-processed by removing those with a length shorter than 300
172 nucleotides. The remaining reads were then processed through QIIME denoiser.

173 Denoised sequence reads were quality controlled and chimeras were detected using
174 UPARSE (Edgar, 2013). Operational taxonomic unit (OTU) sequence representatives
175 were aligned using PyNAST (Caporaso et al., 2010a) with the pre-aligned Greengenes
176 16S core set (DeSantis et al., 2006) as a template and taxonomically classified using
177 Mothur Bayesian classifier (Schloss et al., 2009). The reference database was the SILVA
178 reference database (Pruesse et al., 2007) modified to include sequences from our in-
179 house, curated northern 16S rRNA gene sequence database. Sequences classified as
180 plastid or mitochondrial 16S were removed from the analyses.

181

182 **2.4 Phylogenetic analyses**

183 All phylogenetic analyses were based on a phylogenetic tree constructed with an
184 approximate maximum-likelihood (ML) approach using FastTree v.2.1 (Price et al.,
185 2010) following the procedures described in Monier et al. (2015). UniFrac dw4000
186 (weighted) and duw4000 (unweighted) distances (Lozupone and Knight, 2005) among
187 the different microbial communities were all computed based on the OTU approximate
188 ML phylogenetic tree. Clustering of UniFrac distances was performed using the
189 unweighted pair group method with arithmetic mean (UPGMA) algorithm, and cluster
190 robustness was assessed using 1000 jackknife replicates (on 75% subsets). β -Diversity
191 significance was assessed using UniFrac Monte Carlo significance test on dw4000 with
192 10 000 randomizations, as implemented in QIIME.

193 We investigated community phylogenetic diversity as defined by Faith (1992), along
194 with other diversity metrics such as phylogenetic species richness and evenness (Helmus
195 et al., 2007), using the R package 'picante' v1.5 (Kembel et al., 2010). Community
196 phylogenetic structure was investigated with the calculation of the net relatedness index

197 (NRI) that measures the phylogenetic relatedness for each community. Specifically NRI
198 determines if OTUs are more closely related to co-occurring relatives than expected by
199 chance (Webb et al., 2002).

200

201 **2.5 Statistical analyses**

202 All statistical analyses were carried out using R 3.0.3 (R Core Team, 2014). Abiotic
203 and biotic environmental variables were log-transformed, with the exception of pH
204 (already on a log scale). All analyses were performed on the subsampled dataset (4000
205 sequences per sample) with a total number of 2166 OTUs.

206 Dissimilarities in community composition among the different valleys were visualized
207 using cluster and principal coordinate analyses. A rank abundance plot was generated to
208 identify the bacterial dominants.

209 The taxonomic uniqueness of sites as well as the taxa that contribute the most to these
210 compositional differences were evaluated by means of local contribution to beta-diversity
211 (LCBD; Legendre and De Cáceres, 2013). Differences in LCBD, phylogenetic diversity,
212 species richness and structure across spatial scales were tested using ANOVA followed
213 by Tukey's HSD test and regression models to identify links between site uniqueness and
214 environmental variables.

215 Significant associations between the abundance of bacterial OTUs and the five valleys
216 were further assessed by correlation indices (as a measure of habitat preferences),
217 including the point biserial correlation statistic r_{pb} and its group-equalized value $r.g.$ as
218 defined by De Cáceres and Legendre (2009). Permutation tests (1000 permutations)
219 tested the null hypothesis that the abundance of OTUs in ponds of a given valley was not
220 different from their abundances in ponds located in other valleys. Correction for multi-
221 testing was applied using the method of Benjamini and Hochberg (1995) that controls the
222 false discovery rate and is a less stringent condition than Bonferroni. OTUs that were
223 significantly associated with valleys were submitted to BLASTn search in NCBI
224 GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify the lowest level of
225 classification possible. A heatmap was produced to examine the variability in the
226 ecological preference among the 30 most abundant OTUs.

227

228 2.6 Co-occurrence patterns

229 Co-occurrence analyses were performed using the overall dataset and each of the
230 datasets for the 5 individual valleys. The data were filtered by using only those OTUs
231 with a minimum of 20 reads and that were detected in at least three different ponds. This
232 filtering step removed poorly represented OTUs and reduced the network complexity,
233 resulting in a core community of 294 OTUs.

234 Randomness in co-occurrence of OTUs in the regional and individual valley datasets
235 was tested in a null model using the quasiswap algorithm (Miklós and Podani, 2004) and
236 C-score metric (Stone and Roberts, 1990) under 50000 simulations. SES (standardized
237 effect size) was used as a measure of OTU segregation as described in Heino and
238 Grönroos (2013) in order to determine whether this may relate to the overall
239 environmental heterogeneity, the heterogeneity in biotic and abiotic variables separately,
240 or to specific environmental variables. Environmental heterogeneity was determined
241 using homogenization of group dispersion (Anderson et al., 2006) and defined as the
242 mean distances of ponds to the centroid (central point) of each valley. Analyses were
243 conducted on Euclidean distances on standardized variables and based on 1000
244 permutations. Similarly, the homogenization of group dispersion method was used to
245 determine whether communities among ponds within a given valley were more similar
246 than within other valleys.

247 Network analyses were conducted on the filtered OTU dataset. In addition, a total of 8
248 physicochemical variables (DOC, TP, TN, pH, SUVA₂₅₄, COND: conductivity, T: water
249 temperature, DO: **dissolved** oxygen concentration) and 7 biotic variables (Chl-*a*:
250 phytoplankton biomass, **BA: bacterial abundance, PC: abundance of picocyanobacteria,**
251 **PE: abundance of phototrophic picoeukaryotes, Rot: abundance of rotifers, Clad:**
252 **abundance of cladocerans, Cop: abundance of copepods**) data were also included in the
253 network. For each environmental variable, any missing data were estimated as the mean
254 for the corresponding valley and all data were then normalized by subtracting the mean
255 value for the overall study and dividing by the corresponding standard deviation.

256 To examine associations between the bacterial OTUs and their environment, we
257 analyzed the correlations of the OTUs with each other and with biotic and abiotic
258 variables using the maximal information coefficient (MIC; Reshef et al., 2011). The MIC

259 value indicates the strength of the relationship between two variables and is analogous to
260 R^2 in general linear models. MIC does not provide information on the sign of the
261 association between two nodes, and we therefore extracted the linearity metric (MIC- ρ^2)
262 from the edges of the network, which indicates the type of association: an MIC- ρ^2 value
263 greater than 0.2 implies a strong non-linear association and likely ‘non co-existence’
264 among OTUs (Reshef et al., 2011). Computations were carried out using MINE (Reshef
265 et al., 2011). Following the procedure described in Peura et al. (2015), relationships with
266 $p < 0.05$ were selected to construct networks, which corresponded to a MIC cutoff of 0.44
267 depending on the number of samples in our dataset. Parameters for analysis were set to
268 default, and false discovery rates (Benjamini and Hochberg, 1995) were below 0.03. MIC
269 matrices were translated into networks using Cytoscape 3.2.0 (Shannon et al., 2003).
270 Nodes represented bacterial OTUs as well as both biotic and abiotic variables, which
271 were connected by edges that denote the strength of the relationship between two
272 variables (MIC). The topology of the resulting undirected network was investigated using
273 the package igraph (Csardi and Nepusz, 2006) in R and compared to an Erdős–Rényi
274 random network of similar size. Following Peura et al. (2015), high degree nodes were
275 defined as ‘hubs’ and the implication of their removal for network topology was
276 evaluated. Networks were then visualized in Gephi 0.8.2 (Bastian et al., 2009) using the
277 Fruchterman Reingold layout algorithm. Unconnected nodes were removed along with
278 self-loops and duplicated edges.

279 The relationship between the connectivity of OTUs (as indicated by the degree value
280 in the network) and their corresponding abundance was examined in generalized linear
281 models in order to relax the normality assumptions. OTU abundance was first calculated
282 per individual pond as the product of % of total reads and total bacterial abundance. The
283 total abundance of an OTU in the dataset was then obtained by summing the abundance
284 calculated for each pond. A heatmap was produced to examine the variability in the
285 ecological preference among the 30 most connected OTUs.

286

287 **3 Results**

288 **3.1 Bacterial phylogenetic structure**

289 The phylogenetic composition of bacterial communities differed significantly among

290 valleys (dw4000, UniFrac weighted significance test; $p \leq 0.01$). The clustering and
291 principal coordinate analyses (PCoA) based on weighted UniFrac distances (dw4000;
292 Fig. 1A, 1B) suggested that communities within the SAS valley tend to clustered
293 together, as did the KWK communities. However, a test for homogeneity of multivariate
294 dispersions did not support this as no significant difference in the distance to group
295 (valley) centroid was detected ($P=0.39$, $F=1.08$). Permafrost landscape type had a
296 significant, effect on phylogenetic composition (Permutational analysis of variance on
297 dw4000; $R^2=0.31$, $P=0.001$). The reference lakes did not group together, likely reflecting
298 their disparate catchment properties. The cluster analysis based on unweighted UniFrac
299 distances indicated a stronger clustering according to permafrost landscape type
300 (Permutational analysis of variance on duw4000; $R^2=0.51$; $P=0.001$) by comparison with
301 weighted UniFrac distances (SI Fig. 1; UniFrac unweighted significance test, $p \leq 0.01$).
302 The discrepancy between dw4000 and duw4000 patterns indicated the presence of a
303 small number of highly abundant OTUs within different valleys (SI Fig. 2). In fact, only
304 18 OTUs had a $>1\%$ contribution to the total number of sequence reads.

305 Community phylogenetic analysis based on NRI indices showed that all site clusters
306 had significant phylogenetic structure (positive NRI values; one sample t-test, $t = 18.9$, df
307 $= 33$, $P < 0.0001$; SI Table 1), indicating that bacterial communities within each valley
308 were more closely related to each other than expected by chance. There was no
309 significant difference in phylogenetic structure among valleys (ANOVA, $P=0.4$; Fig.1C),
310 but large differences within individual valleys, with some ponds less phylogenetically
311 clustered than others. For example, the NAS valley two ponds had higher NRI values
312 than the majority of the ponds located within the valley. Ponds located within the SAS
313 valley showed significantly higher phylogenetic species richness and diversity than the
314 KWK, NAS and BGR valleys (PSR: $P=0.002$, $F=5.6$, $R^2=0.36$; PD: $P < 0.0001$, $F=11.3$,
315 $R^2=0.55$).

316 **3.2 Spatial bacterial taxonomic distribution**

317 The local contribution to beta-diversity (LCBD) values indicated the compositional
318 uniqueness of local bacterial communities. One-way ANOVA showed that pond location
319 had a significant influence on compositional uniqueness ($F=2.8$, $R^2=0.27$, $P=0.04$), with
320 the rock basin lakes having the highest LCBD estimates (SI Fig. 3). There was high

321 variability among ponds within the same valley, and there was no significant difference
322 in taxonomic uniqueness among permafrost valleys. Stepwise backward selection
323 identified the best regression model for LCBD as a function of environmental variables
324 (SI Table 2), with four environmental variables ($F=3.2$, $R^2=0.22$, $P=0.03$): DOC,
325 conductivity, $SUVA_{254}$ and Chl-*a*. Sites with a high degree of taxonomic uniqueness had
326 high DOC content and conductivity but low level Chl-*a*. $SUVA_{254}$ made no significant
327 contribution to the model ($P=0.07$), and there was no relationship between LCBD,
328 species richness and distance to the closest neighbor.

329 The thaw pond communities were dominated by OTUs that were assigned to
330 Betaproteobacteria, particularly the order Burkholderiales that was well represented in all
331 communities (35.4% of the total number of reads). Actinobacteria (24.5% of total reads)
332 were mainly represented by OTUs assigned to the family ACK-M1 (60.5% of
333 Actinobacteria reads). Among Bacteroidetes, which accounted for up to 15.7% of the
334 total number of reads, Shingobacteriales were highly represented and were dominated by
335 the family Chitinophagaceae that contributed up to 4.7% of total number of reads. Other
336 dominant OTUs were within the Verrucomicrobia (6.8% of total reads) (Table 1). Among
337 the 30 most abundant taxa, some were highly associated with a specific valley whereas
338 others were not detected in certain valleys (Fig. 2A). This pattern remained when
339 considering the ensemble of the 2166 OTUs (SI Fig 4). Specifically, 272 OTUs (11.3% of
340 the 2166 detected in this dataset) showed a significant association in the indicator value
341 analysis (the point biserial statistic r_g) considering habitat combinations. Among the 272
342 OTUs showing a significant habitat preference, 246 were associated with a single valley:
343 13, 12, 31, 99 and 91 OTUs were associated with the BGR, NAS, KWK, SAS and RBL
344 valleys respectively. Four OTUs were associated with the discontinuous permafrost
345 landscape and three with the sporadic permafrost landscape (Table 2). There were
346 distinctions between ponds located in the sporadic versus discontinuous permafrost
347 landscapes. In particular, OTUs closely related to methanotrophs were prominent within
348 the sporadic permafrost landscape type: OTUs closely related to *Methylothera* (OTU 10)
349 and *Methylobacter* (OTU 9) were among the five most abundant taxa at SAS sites (3.5
350 and 3.6 % of the total number of SAS reads respectively) and OTUs assigned to

351 methanotrophic Verrucomicrobia *LD19* (in the class *Methylacidiphilae*) was one of the
352 most abundant at the KWK site (Fig. 2A, 1.4 % of KWK reads).

353

354 **3.3 Bacterial co-occurrence patterns**

355 To test for differences in co-occurrence patterns between microbial communities
356 across the permafrost landscape, we first selected OTUs that had at least 20 reads and
357 were detected in at least 3 different ponds. The bacterial OTUs were not randomly
358 distributed among the different valleys when considering the entire region (C -score
359 =35.7, $P < 0.0001$, $SES=25.4$). At the individual valley scale, the OTUs were not
360 randomly distributed among ponds except for BGR valley (Table 3). No significant
361 relationship was detected between the level of OTUs segregation, determined by SES,
362 and the overall environmental heterogeneity, and both abiotic and biotic heterogeneity. In
363 addition, no significant relationship between SES and individual environmental variables
364 was detected.

365 The OTU co-occurrence patterns as well as the relationships among both biotic and
366 abiotic variables were investigated by network analysis. The most connected nodes
367 (degree >10) were related to three abiotic variables (DOC, conductivity and TP) and one
368 biotic variable (phototrophic picoeukaryotes). The topology of the networks is presented
369 in Table 4. For the whole regional network, a total of 248 nodes and 968 edges were
370 detected, which was fragmented in 3 components including 2 small components
371 composed of 2 and 3 nodes (SI Fig. 5). The observed characteristic path length of 3.06
372 and clustering coefficient of 0.25 were both greater than estimates originating from the
373 random network of similar size. In addition, the observed:random network clustering
374 coefficient ratio (log response ratio of 0.92) showed that the network had ‘small world’
375 properties; i.e., the nodes were more connected than expected in a random network
376 (Table 4). The frequency distribution of nodes followed a power law function, which
377 indicated that the network was composed of few highly connected nodes, as opposed to
378 an even distribution of connectivity (SI Fig. 6).

379 Four main bacterial phyla were well represented in the networks: Proteobacteria (83
380 nodes), Bacteroidetes (56 nodes), Actinobacteria (42 nodes), and Verrucomicrobia (24
381 nodes). Although edges between nodes that referred to bacterial OTUs dominated the

382 network, connection between bacterial OTUs and both biotic and abiotic variables were
383 detected (SI Fig. 5). For example, conductivity and DOC were amongst the most
384 connected nodes, illustrating their importance in the network. The subnetwork built
385 around DOC showed a diverse bacterial consortium with a slight dominance of
386 Actinobacteria (Fig. 3A). Phototrophic picoeukaryotes were the most connected node
387 among biotic variables. The subnetwork built around that variable showed strong co-
388 occurrence between picoeukaryotes and Actinobacteria (Fig. 3B). The co-occurrence
389 network around the group Chitinophagaceae showed that these OTUs were associated
390 with different environmental variables including DOC, dissolved oxygen, conductivity,
391 abundance of picoeukaryotes, cladocerans and rotifers (Fig. 4A) and had recurrent, strong
392 co-occurrences with Actinobacteria, especially with organisms closely related to ACK-
393 M1 (Fig. 4B). The analysis of the linearity of the latter association indicated a positive
394 co-occurrence between OTUs closely related to members affiliated to the ACK-M1 (aka
395 AcI) group of Actinobacteria and Chitinophagaceae (Fig. 5C). Other examples of strong
396 linkages between OTUs are given in Figure 5, with illustrations of positive co-occurrence
397 (Fig. 5A) and non co-existence (Fig. 5B).

398 In general, our results indicated that the most abundant OTUs were also the most
399 connected ones ($R^2=0.25$, $P<0.001$, SI Fig. 7). However, some of the most connected
400 nodes (OTUs) had low abundance (SI Table 3, Fig. 2B). Noteworthy, some of these
401 bacterial hubs showed some level of habitat preference, especially within KWK valley
402 (Fig. 2B). In addition, these 'valley specific' hubs were mainly related to Actinobacteria
403 and Betaproteobacteria (Fig. 2B).

404 We further investigated the implications of the removal of the top 24 connected OTU
405 nodes (hubs), which represented a removal of 10% of nodes and the results showed a
406 high level of fragmentation of the network and a drop in node degree (Table 4, SI Fig 8).
407 Analysis of the network hubs further showed that the top 24 were mainly composed of
408 Actinobacteria OTUs, in particular members of Actinomycetales and Acidimicrobiales.
409 In addition, OTUs assigned to Betaproteobacteria represented a large fraction of these
410 highly connected OTUs including the typical freshwater *Limnohabitans*, whereas
411 Verrucomicrobia and Bacteroidetes were represented by only a few highly connected

412 OTUs. Interestingly, the anaerobic photosynthetic sulphur bacterium Chloroflexi was also
413 identified as a hub in the overall network (SI Table 3).

414

415 **4 Discussion**

416 The main goal of the present study was to identify co-occurrence patterns among
417 bacterial communities in thaw ponds and lakes in the changing subarctic landscape.
418 Consistent with our first hypothesis, there was a non-random distribution of bacterial taxa
419 across the distinct valleys sampled in this study. The results showed that thaw ponds
420 communities from the same valley, especially those located in the sporadic permafrost
421 landscape, tended to be more similar in terms of bacterial community composition than
422 communities originating from ponds located in other valleys. Furthermore, the thaw
423 ponds differed taxonomically from the rock-basin reference lakes, with specific bacterial
424 OTUs associated with a particular valley or permafrost landscape type. Contrary to our
425 second hypothesis, that differences in habitat preferences among bacterial communities
426 were related to distinct phylogenetic structure, we found no evidence for differences in
427 the community phylogenetic relatedness between the different valleys. The same bacterial
428 phyla occurred throughout the region, and variability among ponds in the same valley
429 was greater than the differences among valleys.

430

431 **4.1 Local community composition uniqueness and habitat preference among** 432 **bacterial communities**

433 Non-random distribution patterns among bacterial taxa were detected, indicating that
434 bacterial taxa in our study region tended to co-occur more than expected by chance. Non-
435 random assembly patterns indicate the dominance of deterministic processes such as
436 environmental filtering in shaping community composition (Horner-Devine et al., 2007).
437 The bacterial communities of freshwater ecosystems elsewhere (Eiler et al., 2011), as
438 well as in certain terrestrial (Barberan et al., 2012) and marine (Steele et al., 2011)
439 ecosystems, have also been reported to have distributional patterns that relate to the
440 environment. Such patterns may depend on niche breadth and competitive abilities
441 (Székely et al., 2013), grazing and viral lysis susceptibilities (Chow et al., 2014; Miki,
442 2008) and dispersal capabilities (Fahlgren et al., 2010; Hervas and Casamayor, 2009).

443 The patterns described here are for the free-living fraction of bacterial assemblages,
444 which raises the question of whether such patterns remain for the attached fraction of the
445 communities. The latter may represent a substantial part of the total communities given
446 that these waterbodies can contain a large content of suspended solids. Previous studies
447 comparing the compositional patterns in bacterial communities between the free-living
448 and attached fractions showed that these two distinct life-style have a similar community
449 composition (Crevecoeur et al., 2015), indicating that the patterns described here may
450 reflect patterns for the entire community.

451 No significant relationship was found between distribution patterns and environmental
452 heterogeneity. This was unexpected, as previous studies have shown that thermokarst
453 systems are heterogeneous environments with marked differences in community
454 composition across the different valleys associated with distinct environmental variables
455 (Crevecoeur et al., 2015; Comte et al., 2015). In agreement with Heino and Grönroos
456 (2013), we suggest that the relationship between distribution pattern and environmental
457 heterogeneity may be scale-dependent such that environmental heterogeneity may have
458 effects on the bacterial taxa distribution patterns at the overall study region scale and not
459 at the valley scale as tested here. The results did show differences in the phylogenetic
460 composition of bacterial communities among the different valleys, which highlight
461 distinct habitat preferences among taxa (Fig. 2, SI Fig. 4). In particular, the combination
462 of LCBD and regression analyses indicated that the compositional uniqueness of thaw
463 ponds and lakes was positively related to DOC concentrations, a well known determinant
464 of bacterial communities and processes (Kritzberg et al., 2006; Ruiz-González et al.,
465 2015). Along with the variations in permafrost degradation state across the study region,
466 there were also differences among valleys in terms of availability and origin of carbon
467 subsidies. The northern sites are located within the discontinuous permafrost area where
468 most of the soil remains frozen and is thus not available for microbial degradation, while
469 in the southern sporadic area, permafrost is highly degraded (Bouchard et al., 2014) and
470 large amounts of ancient permafrost carbon may be available for microbial processes.
471 Consistent with this pattern, elevated concentrations and high rates of CO₂ and CH₄
472 emission to the atmosphere have been reported among the southern sites within the most
473 degraded area of permafrost (Laurion et al., 2010; Deshpande et al., 2015). This may in

474 turn explain the significantly higher bacterial richness and diversity observed in SAS
475 thaw ponds communities and why OTUs assigned to methanotrophic bacteria such as
476 *Methylobacter* and *Methylothera* were amongst the most abundant detected in this valley
477 (Fig. 2). In addition, SAS sites originated from palsas (organic permafrost mounds) and
478 were likely different in DOC composition relative to other valleys, where the ponds were
479 formed by the thawing of lithalsas (mineral permafrost mounds). This is consistent with
480 recent observation of a direct link between community composition and the degradation
481 of terrestrially derived DOM (Logue et al., 2015).

482

483 **4.2 Bacterial phylogenetic structure**

484 The mean NRI across all communities was significantly greater than zero. This
485 provides evidence for a dominant role of environmental filtering on community
486 composition (Kembel, 2009). The corollary is that a set of environmental variables
487 constrained community composition, resulting in taxa that were closer phylogenetically
488 and more ecologically similar than if stochastic processes (including dispersal) drove
489 community assembly. In fact, there is no corridor such as streams that connects the
490 ponds, and thus local dispersal processes are unlikely to explain the local phylogenetic
491 structure of the thaw pond communities (Comte et al., 2015). Similar results were
492 obtained for microbial community studies in the ocean (Monier et al., 2015) and on
493 groundwater communities (Stegen et al., 2012).

494 No significant difference in NRI was found among the different valleys, but this result
495 likely reflects the high variability within individual valleys. In particular, two ponds in
496 the NAS valley had higher values of NRI in comparison to their neighboring ponds.
497 These two ponds had specific environmental characteristics including high concentrations
498 of suspended clay particles and low phytoplankton concentrations, which may have
499 favored certain environmental specialists. The rock-basin waters had higher NRI values
500 than the thaw ponds, indicating that their assemblages were more ecologically similar to
501 each other than those originating from thaw ponds and lakes. This could relate to their
502 respective histories in that the rock- basin lakes originate from deglaciation followed by
503 retreat of the Tyrell Sea ca. 8000 years ago and have thus been exposed to longer term
504 ecological processes.

505 The extent of permafrost erosion (permafrost landscape type) appeared to influence
506 phylogenetic structure. When controlling for the two outliers mentioned above (NAS-A
507 and NAS-B), the northern communities (BGR, NAS) had a greater phylogenetic distance
508 among co-occurring taxa than expected by chance (lower NRIs) in comparison to
509 communities from the thaw ponds located in valleys from sporadic permafrost (KWK,
510 SAS). This suggests that taxa from SAS valley (and to a lesser extent KWK), tend to be
511 more ecologically similar to each other than those from northern valleys, reflecting strong
512 environmental filtering by variables such as DOC concentration, as previously
513 documented in this valley (Comte et al., 2015). These findings are in line with studies
514 elsewhere that showed that clustered communities are mainly retrieved from
515 environments that have constrained environmental conditions (Monier et al., 2015).

516

517 **4.3 Network associations**

518 The extent to which closely related bacterial taxa may coexist is still a subject of
519 considerable discussion (Mayfield and Levine, 2010). Previous studies on aquatic
520 microbial communities have shown that closely related taxa have coherent temporal
521 dynamics and share similar ecological niches (Andersson et al., 2009; Eiler et al., 2011).
522 Co-occurrence networks enable the depiction and visualization of co-occurrence patterns
523 among OTUs, and they provide a way of identifying potential ecological niches within
524 microbial consortia. Network analyses have recently been applied to a wide range of
525 microbial communities and biomes, and specific associations among bacterial OTUs and
526 with environmental variables have been reported (Barberan et al., 2012; Chow et al.,
527 2014; Eiler et al., 2011; Steele et al., 2011).

528 Our results point toward the importance of environmental filtering for community
529 assembly in thaw ponds and lakes. In co-occurrence networks, correlations between
530 OTUs and environmental variables highlight the conditions that may favor particular
531 assemblages. Specifically, our co-occurrence networks identified two abiotic variables
532 (DOC and conductivity) to be among the most connected nodes (SI Fig. 5B), and these
533 variables separated according to landscape type: the northern ponds located in the
534 discontinuous permafrost landscape had high conductivity and low DOC, whereas
535 southern sites within the sporadic permafrost landscape had high DOC and lower

536 conductivity (SI Table 2; further details are given in Comte et al., 2015). The analysis of
537 the DOC subnetwork showed that only a few OTUs were significantly and directly
538 related to DOC; these included OTUs assigned to Actinobacteria as well as OTUs closely
539 related to bacterial methanotrophs and taxa involved in the degradation of complex
540 organic polymers (Fig. 3A). Among phylogenetically related microbes, unique
541 combinations tended to co-occur (Fig. 4A). For example, some OTUs assigned to the
542 Chitinophagaceae appeared to be significantly related to different abiotic and biotic
543 variables, which in turn suggested niche separation.

544 In addition to the bottom-up factors that shape bacterial communities, recent work on
545 microbial networks has highlighted the role of top down processes such as grazing and
546 viral lysis in affecting prokaryotic community structure and co-occurrence patterns
547 (Chow et al., 2014; Steele et al., 2011). In the present study, phototrophic picoeukaryote
548 abundance (degree=14) was the most connected biotic node. Only phototrophic
549 picoeukaryotes were enumerated in this study, and although some may have a
550 mixotrophic grazing capacity, their network importance may be the result of other
551 factors, for example the release of photosynthate or their occurrence under conditions that
552 mutually favor both themselves and certain bacterial taxa.

553 In general, relationships among microbes dominated the network, rather than those
554 between microbes and abiotic or biotic environmental parameters (SI Fig.5). There was
555 overlap in terms of community composition among the different valleys (Fig 1), with
556 shared dominant taxa (Table 1, SI Fig. 2). Although this may indicate that some OTUs
557 may respond similarly to specific environmental factors and outcompete others, some
558 associations may be the result of substrate interdependencies. One example is the
559 relationship between bacteria able to degrade chitin and others that take up the resulting
560 hydrolysis products (Beier and Bertilsson, 2013). OTUs closely related to bacteria in the
561 Chitinophagaceae, a group known to be involved in the degradation of chitin and other
562 complex polymeric organic matter (del Rio et al., 2010), were well represented in our
563 study area, and have also been found in other cold terrestrial environments (Franzetti et
564 al., 2013; Ganzert et al., 2011). The subnetwork built around this group showed that these
565 OTUs are linked to other phyla (Fig. 4A), notably certain Actinobacteria (Fig. 4B). The
566 dominants were closely related to clade Ac1, which is known to include specialists that

567 use hydrolysis products from chitinolytic bacteria (Beier and Bertilsson, 2011). The
568 analysis of linearity of the associations between the corresponding OTUs showed a
569 positive co-occurrence (Fig. 5C), consistent with bacterial network relationships.
570 Although other examples of positive co-occurrence among bacterial OTUs were
571 identified in the dataset (Fig. 5A), there was also evidence of ‘non co-existence’ (*sensu*
572 *Reshef et al., 2011*) among certain OTUs: In the northern, less degraded permafrost
573 valley (BGR), OTU 1242 (Betaproteobacteria *Limnohabitans*) dominated, whereas in the
574 southern **highly** degraded permafrost valleys (SAS, KWK), OTU 14 (Actinobacteria
575 *ACK-M1*) dominated (Fig. 5B). These trade-offs among OTUs were partially explained
576 by the geographic location of the valleys, suggesting that environmental variables not
577 only drive the composition of the bacterial assemblages within the individual valleys but
578 may also determine the ecological associations within microbial consortia. Furthermore,
579 the positive relationship found between the connectivity and the habitat specificity among
580 the most abundant OTUs (Fig. 2A) is most likely driven by the dominance of highly
581 connected OTUs in the southern highly degraded permafrost valleys in comparison to the
582 northern less degraded permafrost valleys. In addition, the OTUs retrieved from the
583 southern thaw ponds were closely related to specific bacterial functional groups such as
584 methanotrophs and nitrogen fixing bacteria (Fig. 5).

585

586 The microbial networks for the thermokarst systems had ‘small world’ properties, with
587 only a few, highly connected nodes, which can be viewed as ‘keystone species’. This
588 property would render the networks more resilient to environmental change, but
589 vulnerable to the loss of these nodal species (Montoya et al., 2006). The bacterial hubs
590 were identified as typical freshwater, terrestrial and marine taxa (SI Table 3), and some of
591 them were closely related to taxa that are involved in key biogeochemical processes such
592 as nitrogen fixation and degradation of complex polymers, or that are known to be
593 restricted in niche breadth, for example to cold environments. In accordance with Peura et
594 al. (2015), the importance of a taxon in a microbial network may be less associated with
595 its abundance, but instead determined by its connectivity, as represented by node degree
596 for example. Thus many of the hub taxa identified in this study could be defined as a
597 keystone microbial species (SI Table 3). These ‘keystone’ OTUs identified as hubs were

598 not merely the abundant OTUs (Fig. 2B), but some were rare and potentially important
599 actors for the functioning of these ecosystems. For example, the nitrogen-fixing
600 bacterium *Beijerinckia* was among the most connected node in the co-occurrence network
601 despite its low relative abundance. **This in turn highlights the potentially important
602 ecological role of diazotrophs in these nutrient-rich aquatic systems.**

603

604 **Conclusions**

605 The thaw ponds and lakes sampled in the present study showed large variability in
606 their bacterial community structure, even among **waterbodies** in a single valley. This
607 underscores the heterogeneous nature of permafrost aquatic environments, and is
608 consistent with their known limnological variability. A small number of taxa occurred in
609 high abundance and dominated many of the communities; these northern dominants
610 included members of the betaproteobacterial order Burkholdiales and the Actinobacterial
611 family ACK-M1; other dominants included members of the Bacteroidetes family
612 Chitinophagaceae and Verrucomicrobia. Despite this variability and the existence of
613 common taxa, there were taxonomic differences among different valleys and between
614 permafrost landscape types, implying some degree of habitat selection.

615 The bacterial networks further showed that DOC and conductivity played an important
616 role in the co-occurrence patterns of bacterial OTUs, corresponding at least in part to
617 differences in these two environmental variables among valleys (SI Table 2). Strong
618 positive associations as well as non-coexistence among OTUs were detected, and the
619 resultant networks were composed of a limited number of highly connected OTUs. This
620 ‘small world network’ property would render these communities more resilient to
621 environmental change, but sensitive to the loss of their hub OTUs, which themselves
622 showed some degree of habitat specificity. With ongoing global warming, these waters
623 are likely to experience the effects of increased permafrost erosion and associated
624 changes in their chemical environment, including shifts in DOC and conductivity. If such
625 changes eventually cause the loss of ‘keystone species’ that form the hubs of the present
626 microbial networks, there would be a major disruption of thaw ponds and lakes
627 community structure, with potentially large biogeochemical consequences.

628

629 *Acknowledgements*

630 We are grateful to M. Bartosiewicz, B. Deshpande, A. Matveev, A. Przytulska-
631 Bartosiewicz as well as C. Tremblay from Whapmagoostui-Kuujjuarapik CEN station
632 and the pilots of Canadian Helicopter Ltd., for their assistance in the field. We are also
633 grateful to Paschale N. Begin for zooplankton enumeration, **M.-J. Martineau for pigment**
634 **analyses, I. Laurion (INRS-ETE) for flow cytometry.** Computing support from
635 CLUMEQ/Compute Canada, aid from A. Monier for bioinformatics and phylogenetic
636 analyses, advice from A. Eiler for network analyses, and insightful comments from two
637 anonymous reviewers **and the Editor were also greatly appreciated.** We acknowledge the
638 Natural Sciences and Engineering Council (NSERC) of Canada funding for Discovery
639 grants to WFV and CL and Discovery Frontier (ADAPT) grant to WFV, the support from
640 the Network of Centres of Excellence program ArcticNet to WFV and CL, and the
641 Canadian Research Chair Program to WFV. Additional support from Fonds de Recherche
642 du Québec Nature et Technologies (FRQNT) to CEN is acknowledged. JC was partially
643 supported by a FRQNT postdoctoral fellowship and the EnviroNorth CREATE program
644 from NSERC.

645

646 **References**

647 Abnizova, A., Siemens, J., Langer, M. and Boike, J.: Small ponds with major impact: The
648 relevance of ponds and lakes in permafrost landscapes to carbon dioxide emissions,
649 *Global Biogeochem. Cy.*, 26(2), 2012.

650 Anderson, M. J., Ellingsen, K. E. and McArdle, B. H.: Multivariate dispersion as a
651 measure of beta diversity, *Ecol. Lett.*, 9(6), 683–693, doi:10.1111/j.1461-
652 0248.2006.00926.x, 2006.

653 Andersson, A. F., Riemann, L. and Bertilsson, S.: Pyrosequencing reveals contrasting
654 seasonal dynamics of taxa within Baltic Sea bacterioplankton communities, *ISME J*, 4(2),
655 171–181, doi:10.1038/ismej.2009.108, 2009.

656 Andresen, C. G. and Lougheed, V. L.: Disappearing Arctic tundra ponds: Fine-scale
657 analysis of surface hydrology in drained thaw lake basins over a 65 year period (1948-
658 2013), *J. Geophys. Res. Biogeosci.*, 120, doi:10.1002/2014JG002778, 2015.

659 Araújo, M. B., Rozenfeld, A., Rahbek, C. and Marquet, P. A.: Using species co-
660 occurrence networks to assess the impacts of climate change, *Ecography*, 34(6), 897–908,
661 2011.

- 662 Barberan, A., Bates, S. T., Casamayor, E. O. and Fierer, N.: Using network analysis to
663 explore co-occurrence patterns in soil microbial communities, *ISME J*, 6(2), 343–351,
664 doi:10.1038/ismej.2011.119, 2012.
- 665 Bastian, M., Heymann, S. and Jacomy, M.: Gephi: an open source software for exploring
666 and manipulating networks, *ICWSM*, 8, 361–362, 2009.
- 667 Beier, S. and Bertilsson, S.: Uncoupling of chitinase activity and uptake of hydrolysis
668 products in freshwater bacterioplankton, *Limnol. Oceanogr.*, 56(4), 1179–1188,
669 doi:10.4319/lo.2011.56.4.1179, 2011.
- 670 Beier, S. and Bertilsson, S.: Bacterial chitin degradation-mechanisms and
671 ecophysiological strategies, *Front. Microbiol.*, 4, 149, doi:10.3389/fmicb.2013.00149,
672 2013.
- 673 Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and
674 powerful approach to multiple testing, *J. Roy. Stat. Soc. B*, 289–300, 1995.
- 675 Bhiry, N., Delwaide, A., Allard, M., Bégin, Y., Filion, L., Lavoie, M., Nozais, C.,
676 Payette, S., Pienitz, R., Saulnier-Talbot, É. and Vincent, W. F. : Environmental change in
677 the Great Whale River region, Hudson Bay: Five decades of multidisciplinary research by
678 Centre d'études nordiques (CEN), *Ecoscience* 18, 182–203, 2011.
- 679 Bonilla, S., Villeneuve, V. and Vincent, W. F.: Benthic and planktonic algal communities
680 in a high arctic lake: Pigment structure and contrasting responses to nutrient enrichment,
681 *J. Phycol.*, 41(6), 1120–1130, 2005.
- 682 Bouchard, F., Francus, P., Pienitz, R., Laurion, I. and Feyte, S.: Subarctic thermokarst
683 ponds: Investigating recent landscape evolution and sediment dynamics in thawed
684 permafrost of northern Québec (Canada), *Arct. Antarct. Alp. Res.*, 46(1), 251–271,
685 doi:10.1657/1938-4246-46.1.251, 2014.
- 686 Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. and
687 Knight, R.: PyNAST: a flexible tool for aligning sequences to a template alignment,
688 *Bioinformatics*, 26(2), 266–267, doi:10.1093/bioinformatics/btp636, 2010a.
- 689 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E.
690 K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., et al.: QIIME
691 allows analysis of high-throughput community sequencing data, *Nat. Methods*, 7(5), 335–
692 336, doi:10.1038/nmeth.f.303, 2010b.
- 693 Chow, C.-E. T., Kim, D. Y., Sachdeva, R., Caron, D. A. and Fuhrman, J. A.: Top-down
694 controls on bacterial community structure: microbial network analysis of bacteria, T4-like
695 viruses and protists, *ISME J*, 8(4), 816–829, doi:10.1038/ismej.2013.199, 2014.
- 696 Comeau, A. M., Li, W. K. W., Tremblay, J.-É., Carmack, E. C. and Lovejoy, C.: Arctic
697 Ocean microbial community structure before and after the 2007 record sea ice minimum,
698 *PLoS ONE*, 6(11), e27492, doi:10.1371/journal.pone.0027492.s012, 2011.

- 699 Comte, J., Monier, A., Crevecoeur, S., Lovejoy, C. and Vincent, W. F.: Microbial
700 biogeography of permafrost thaw ponds across the changing northern landscape,
701 *Ecography*, 38, doi: 10.1111/ecog.01667, 2015.
- 702 Crevecoeur, S., Vincent, W. F., Comte, J. and Lovejoy, C.: Bacterial community structure
703 across environmental gradients in permafrost thaw ponds: methanotroph-rich ecosystems,
704 *Front Microbiol*, 6, 192, doi:10.3389/fmicb.2015.00192, 2015.
- 705 Csardi, G. and Nepusz, T.: The igraph software package for complex network research,
706 *InterJ. Complex Sys.*, 1695(5), 1–9, 2006.
- 707 De Cáceres, M. and Legendre, P.: Associations between species and groups of sites:
708 indices and statistical inference, *Ecology*, 90(12), 3566–3574, 2009.
- 709 del Rio, T. G., Abt, B., Spring, S., Lapidus, A., Nolan, M., Tice, H., Copeland, A.,
710 Cheng, J.-F., Chen, F., Bruce, D., Goodwin, L., et al.: Complete genome sequence of
711 *Chitinophaga pinensis* type strain (UQM 2034), *Stand. Genomic Sci.*, 2(1), 87–95,
712 doi:10.4056/sigs.661199, 2010.
- 713 DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber,
714 T., Dalevi, D., Hu, P. and Andersen, G. L.: Greengenes, a chimera-checked 16S rRNA
715 gene database and workbench compatible with ARB, *Appl. Environ. Microbiol.*, 72(7),
716 5069–5072, doi:10.1128/AEM.03006-05, 2006.
- 717 Deshpande, B., MacIntyre, S., Matveev, A., and Vincent, W. F.: Oxygen dynamics in
718 permafrost thaw lakes: Anaerobic bioreactors in the Canadian subarctic, *Limnol.*
719 *Oceanogr.* 60(5), 1656–1670, doi: 10.1002/lno.10126, 2015.
- 720 Ducklow, H.: Microbial services: challenges for microbial ecologists in a changing
721 world, *Aquat. Microb. Ecol.*, 53, 13–19, doi:10.3354/ame01220, 2008.
- 722 Edgar, R. C.: UPARSE: highly accurate OTU sequences from microbial amplicon reads,
723 *Nat. Methods*, 10(10), 996–998, doi:10.1038/nmeth.2604, 2013.
- 724 Eiler, A., Heinrich, F. and Bertilsson, S.: Coherent dynamics and association networks
725 among lake bacterioplankton taxa, *ISME J*, 6(2), 330–342, doi:10.1038/ismej.2011.113,
726 2011.
- 727 Fahlgren, C., Hagström, A., Nilsson, D. and Zweifel, U. L.: Annual variations in the
728 diversity, viability, and origin of airborne bacteria, *Appl. Environ. Microbiol.*, 76(9),
729 3015–3025, doi:10.1128/AEM.02092-09, 2010.
- 730 Faith, D. P.: Conservation evaluation and phylogenetic diversity, *Biol. Conserv.*, 61(1),
731 1–10, 1992.
- 732 Franzetti, A., Tatangelo, V., Gandolfi, I., Bertolini, V., Bestetti, G., Diolaiuti, G.,
733 D'Agata, C., Mihalcea, C., Smiraglia, C. and Ambrosini, R.: Bacterial community
734 structure on two alpine debris-covered glaciers and biogeography of *Polaromonas*

- 735 phylotypes, *ISME J*, 7(8), 1483–1492, doi:10.1038/ismej.2013.48, 2013.
- 736 Ganzert, L., Lipski, A., Hubberten, H.-W. and Wagner, D.: The impact of different soil
737 parameters on the community structure of dominant bacteria from nine different soils
738 located on Livingston Island, South Shetland Archipelago, Antarctica, *FEMS Microbiol.*
739 *Ecol.*, 76(3), 476–491, doi:10.1111/j.1574-6941.2011.01068.x, 2011.
- 740 Grosse, G., Jones, B. and Arp, C.: Thermokarst Lakes, Drainage, and Drained Basins. In:
741 John F. Shroder (ed.) *Treatise on Geomorphology*, Volume 8, pp. 325-353. San Diego:
742 Academic Press., 2013.
- 743 Heino, J. and Grönroos, M.: Does environmental heterogeneity affect species co-
744 occurrence in ecological guilds across stream macroinvertebrate metacommunities?
745 *Ecography*, 36(8), 926–936, doi:10.1111/j.1600-0587.2012.00057.x, 2013.
- 746 Helmus, M. R., Bland, T. J., Williams, C. K. and Ives, A. R.: Phylogenetic Measures of
747 Biodiversity, *Am. Nat.*, 169(3), E68–E83, doi:10.1086/511334, 2007.
- 748 Hervàs, A. and Casamayor, E. O.: High similarity between bacterioneuston and airborne
749 bacterial community compositions in a high mountain lake area, *FEMS Microbiol. Ecol.*,
750 67(2), 219–228, doi:10.1111/j.1574-6941.2008.00617.x, 2009.
- 751 Horner-Devine, M. C., Silver, J. M., Leibold, M. A., Bohannon, B. J., Colwell, R. K.,
752 Fuhrman, J. A., Green, J. L., Kuske, C. R., Martiny, J. B. and Muzer, G.: A comparison
753 of taxon co-occurrence patterns for macro- and microorganisms, *Ecology*, 88(6), 1345–
754 1353, 2007.
- 755 Kembel, S. W.: Disentangling niche and neutral influences on community assembly:
756 assessing the performance of community phylogenetic structure tests, *Ecol. Lett.*, 12(9),
757 949–960, doi:10.1111/j.1461-0248.2009.01354.x, 2009.
- 758 Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D.
759 D., Blomberg, S. P. and Webb, C. O.: Picante: R tools for integrating phylogenies and
760 ecology, *Bioinformatics*, 26(11), 1463–1464, doi:10.1093/bioinformatics/btq166, 2010.
- 761 Kritzberg, E. S., Langenheder, S. and Lindström, E. S.: Influence of dissolved organic
762 matter source on lake bacterioplankton structure and function--implications for seasonal
763 dynamics of community composition, *FEMS Microbiol. Ecol.*, 56(3), 406–417,
764 doi:10.1111/j.1574-6941.2006.00084.x, 2006.
- 765 Laurion, I., Vincent, W. F., MacIntyre, S., Retamal, L., Dupont, C., Francus, P. and
766 Pienitz, R.: Variability in greenhouse gas emissions from permafrost thaw ponds, *Limnol.*
767 *Oceanogr.*, 55(1), 115, doi:10.4319/lo.2010.55.1.0115, 2010.
- 768 Legendre, P. and De Cáceres, M.: Beta diversity as the variance of community data:
769 dissimilarity coefficients and partitioning, *Ecol. Lett.*, 16(8), 951–963,
770 doi:10.1111/ele.12141, 2013.

771 Logue, J. B., Stedmon, C. A., Kellerman, A. M., Nielsen, N. J., Andersson, A. F.,
772 Laudon, H., Lindström, E. S. and Kritzberg, E. S.: Experimental insights into the
773 importance of aquatic bacterial community composition to the degradation of dissolved
774 organic matter, *ISME J.*, doi:10.1038/ismej.2015.131, 2015.

775 Lozupone, C. and Knight, R.: UniFrac: a new phylogenetic method for comparing
776 microbial communities, *Appl. Environ. Microbiol.*, 71(12), 8228–8235,
777 doi:10.1128/AEM.71.12.8228-8235.2005, 2005.

778 Mayfield, M. M. and Levine, J. M.: Opposing effects of competitive exclusion on the
779 phylogenetic structure of communities, *Ecol. Lett.*, 13(9), 1085–1093,
780 doi:10.1111/j.1461-0248.2010.01509.x, 2010.

781 McGuire, A. D., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D.
782 J., Heimann, M., Lorensen, T. D., Macdonald, R. W. and Roulet, N.: Sensitivity of the
783 carbon cycle in the Arctic to climate change, *Ecol. Monogr.*, 79(4), 523–555,
784 doi:10.1890/08-2025.1, 2009.

785 Miki, T.: A new graphical model for untangling complex relationships among
786 environment, biodiversity, and ecosystem functioning, *Ecol. Res.*, 24(4), 937–941,
787 doi:10.1007/s11284-008-0552-7, 2008.

788 Miklós, I. and Podani, J.: Randomization of presence-absence matrices: comments and
789 new algorithms, *Ecology*, 85(1), 86–92, 2004.

790 Mondav, R., Ben J Woodcroft, Kim, E.-H., McCalley, C. K., Hodgkins, S. B., Crill, P.
791 M., Chanton, J., Hurst, G. B., VerBerkmoes, N. C., Saleska, S. R., Hugenholtz, P., et al.:
792 Discovery of a novel methanogen prevalent in thawing permafrost, *Nat. Commun.*, 5, 1–
793 7, doi:10.1038/ncomms4212, 2014.

794 Monier, A., Comte, J., Babin, M., Forest, A., Matsuoka, A. and Lovejoy, C.:
795 Oceanographic structure drives the assembly processes of microbial eukaryotic
796 communities, *ISME J.*, 9(4), 990–1002, doi:10.1038/ismej.2014.197, 2015.

797 Montoya, J. M., Pimm, S. L. and Solé, R. V.: Ecological networks and their fragility,
798 *Nature*, 442(7100), 259–264, doi:10.1038/nature04927, 2006.

799 Peura, S., Bertilsson, S., Jones, R. I. and Eiler, A.: Resistant microbial co-occurrence
800 patterns inferred by network topology, *Appl. Environ. Microbiol.*, 81(6), 2090-2097,
801 doi:10.1128/AEM.03660-14, 2015.

802 Price, M. N., Dehal, P. S. and Arkin, A. P.: FastTree 2--approximately maximum-
803 likelihood trees for large alignments, *PLoS ONE*, 5(3), e9490,
804 doi:10.1371/journal.pone.0009490, 2010.

805 R Core Team: R: A language and environment for statistical computing., edited by R
806 Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, 2014.

- 807 Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh,
808 P. J., Lander, E. S., Mitzenmacher, M. and Sabeti, P. C.: Detecting novel associations in
809 large data sets, *Science*, 334(6062), 1518–1524, doi:10.1126/science.1205438, 2011.
- 810 Rossi, P. G., Laurion, I. and Lovejoy, C.: Distribution and identity of bacteria in subarctic
811 permafrost thaw ponds, *Aquat. Microb. Ecol.*, 69(3), 231–245, doi:10.3354/ame01634,
812 2013.
- 813 Ruiz-González, C., Niño-García, J. P., Lapierre, J.-F. and Del Giorgio, P. A.: The quality
814 of organic matter shapes the functional biogeography of bacterioplankton across boreal
815 freshwater ecosystems, *Global Ecology and Biogeography*, in press,
816 doi:10.1111/geb.12356, 2015.
- 817 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B.,
818 Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., et al.:
819 Introducing mothur: Open-source, platform-independent, community-supported software
820 for describing and comparing microbial communities, *Appl. Environ. Microbiol.*, 75(23),
821 7537–7541, doi:10.1128/AEM.01541-09, 2009.
- 822 Schuur, E. A., Bockheim, J., Canadell, J. G., Euskirchen, E., Field, C. B., Goryachkin, S.
823 V., Hagemann, S., Kuhry, P., Lafleur, P. M. and Lee, H.: Vulnerability of permafrost
824 carbon to climate change: Implications for the global carbon cycle, *BioScience*, 58(8),
825 701–714, doi:10.1641/B580807, 2008.
- 826 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N.,
827 Schwikowski, B. and Ideker, T.: Cytoscape: a software environment for integrated
828 models of biomolecular interaction networks, *Genome Res.*, 13(11), 2498–2504, 2003.
- 829 Stainton, M.P., Capel, M. J., and Armstrong, F. A. J.: *The Chemical Analysis of Fresh*
830 *Water*. Winnipeg: Canadian Fisheries and Marine Service. Special Publication 25, 1–168.
- 831 Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., Chow, C.-
832 E. T., Sachdeva, R., Jones, A. C., Schwalbach, M. S., Rose, J. M., et al.: Marine bacterial,
833 archaeal and protistan association networks reveal ecological linkages, *ISME J*, 5(9),
834 1414–1425, doi:10.1038/ismej.2011.24, 2011.
- 835 Stegen, J. C., Lin, X., Konopka, A. E. and Fredrickson, J. K.: Stochastic and deterministic
836 assembly processes in subsurface microbial communities, *ISME J*, 6(9), 1653–1664,
837 doi:10.1038/ismej.2012.22, 2012.
- 838 Stone, L. and Roberts, A.: The checkerboard score and species distributions, *Oecologia*,
839 85(1), 74–79, 1990.
- 840 Székely, A. J., Berga, M. and Langenheder, S.: Mechanisms determining the fate of
841 dispersed bacterial communities in new environments, *ISME J*, 7(1), 61–71,
842 doi:10.1038/ismej.2012.80, 2013.
- 843 Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G. and Zimov, S.:

- 844 Soil organic carbon pools in the northern circumpolar permafrost region, *Global*
845 *Biogeochem. Cy.*, 23(2), n/a–n/a, doi:10.1029/2008GB003327, 2009.
- 846 Thibault, S. and Payette, S.: Recent permafrost degradation in bogs of the James Bay
847 area, northern Quebec, Canada, *Permafrost Periglac.*, 20(4), 383–389, 2009.
- 848 Utermöhl, H.: Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. *Mitt. Int.*
849 *Ver. Theor. Angew. Limnol.* 9, 1-38, 1958.
- 850 Walter Anthony, K. M., Zimov, S. A., Grosse, G., Jones, M. C., Anthony, P. M., Chapin,
851 F. S., Finlay, J. C., Mack, M. C., Davydov, S., Frenzel, P. and Frohling, S.: A shift of
852 thermokarst lakes from carbon sources to sinks during the Holocene epoch, *Nature*,
853 511(7510), 452–456, doi:10.1038/nature13560, 2014.
- 854 Walter, K. M., Chanton, J. P., Chapin, F. S., Schuur, E. and Zimov, S. A.: Methane
855 production and bubble emissions from arctic lakes: Isotopic implications for source
856 pathways and ages, *J. Geophys. Res-Bioge.* (2005–2012), 113(G3), 2008.
- 857 Walter, K. M., Smith, L. C. and Stuart Chapin, F.: Methane bubbling from northern lakes:
858 present and future contributions to the global methane budget, *Philos. Trans. Roy. Soc. A*,
859 365(1856), 1657–1676, doi:10.1126/science.1128908, 2007.
- 860 Webb, C. O., Ackerly, D. D., McPeck, M. A. and Donoghue, M. J.: Phylogenies and
861 community ecology, *Annu. Rev. Ecol. Syst.*, 33(1), 475–505,
862 doi:10.1146/annurev.ecolsys.33.010802.150448, 2002.
- 863 Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R. and Mopper, K.:
864 Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition
865 and reactivity of dissolved organic carbon, *Environ. Sci. Technol.*, 37(20), 4702–4708,
866 doi:10.1021/es030360x, 2003.
- 867
- 868
- 869

Table 1: Five most abundant (number of reads) OTUs across spatial scales. Finest taxonomy assignments are presented with a minimum confidence of 0.8.

All sites	Geographic Categories						
	Landscapes		Valleys				
	<i>Discontinuous</i> (<i>BGR+NAS</i>)	<i>Sporadic</i> (<i>SAS+KWK</i>)	<i>BGR</i>	<i>NAS</i>	<i>KWK</i>	<i>SAS</i>	<i>RBL</i>
<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> ACK_M1	<i>Polynucleobacter</i> Comamonadaceae	Comamonadaceae <i>Polynucleobacter</i>
Comamonadaceae <i>Flavobacterium</i>	Comamonadaceae <i>Flavobacterium</i>	Comamonadaceae ACK_M1	Comamonadaceae <i>Flavobacterium</i>	Comamonadaceae Comamonadaceae	ACK_M1	<i>Polynucleobacter</i> <i>Methylobacter</i>	ACK_M1 Burkholderiales
<i>Arcicella</i>	<i>Arcicella</i>	ACK_M1	<i>Arcicella</i>	<i>Arcicella</i>	<i>Limnohabitans</i>	<i>Methylothera</i>	<i>Arcicella</i>

Table 2: Results of indicator species analysis. Valley refers to the valley (or combination of valleys) for which the OTU obtained the highest correlation. We indicate the correlation value (r.g) and its statistical significance (P) at $\alpha=0.05$. Only OTUs with $r.g \geq 0.6$ are presented when associated to one valley (top 10 are presented for the KWK and SAS valleys). OTUs were classified at their finest taxonomic levels based on similarity to sequences in Genbank.

OTUs	Valley	r.g	P	OTUs	Valley	r.g	P
<u>Discontinuous permafrost</u>				<u>Sporadic Permafrost</u>			
<i>Alterococcus</i>	BGR	0.78	0.004	Oxalobacteraceae	KWK	0.81	0.010
<i>Pseudoclavibacter</i>	BGR	0.75	0.004	<i>Candidatus Planktoluna</i>	KWK	0.80	0.012
<i>Variovorax</i>	BGR	0.71	0.004	Actinomycetales	KWK	0.79	0.010
<i>Alterococcus</i>	BGR	0.65	0.012	Opitutae	KWK	0.74	0.010
<i>Leifsonia</i>	BGR	0.63	0.012	Gammaproteobacteria	KWK	0.67	0.013
<i>Candidatus Protochlamydia</i>	BGR	0.62	0.009	<i>Lacibacter</i>	KWK	0.67	0.012
Thermodesulfobacteriaceae	NAS	0.69	0.012	<i>Burkholderia</i>	KWK	0.64	0.013
<i>Methylosinus</i>	NAS	0.67	0.012	Unknown Proteobacteria	KWK	0.62	0.024
<i>Flavobacterium</i>	NAS	0.67	0.012	Alphaproteobacteria	KWK	0.61	0.024
<i>Ferruginibacter</i>	NAS	0.64	0.012	<i>Mycobacterium</i>	KWK	0.60	0.019
<i>Klugiella</i>	NAS	0.6	0.024	<i>Polynucleobacter</i>	SAS	0.86	0.005
<i>Sporichthya</i>	BGR+NAS	0.59	0.036	Flavobacteriaceae	SAS	0.84	0.005
<i>Arcicella</i>	BGR+NAS	0.53	0.036	<i>Caenimonas</i>	SAS	0.84	0.005
Microbacteriaceae	BGR+NAS	0.51	0.036	Firmicutes	SAS	0.82	0.005
<i>Ferruginibacter</i>	BGR+NAS	0.50	0.043	<i>Polynucleobacter</i>	SAS	0.82	0.005
<u>Rock basin lakes</u>				Alphaproteobacteria	SAS	0.81	0.005
<i>Sphingobium</i>	RBL	0.85	0.011	Anaeomyxobacter	SAS	0.80	0.005
<i>Bordetella</i>	RBL	0.78	0.011	<i>Unclassified bacteria</i>	SAS	0.80	0.005
<i>Neochlamydia</i>	RBL	0.74	0.011	<i>Flavobacterium</i>	SAS	0.80	0.005
<i>Wolbachia</i>	RBL	0.74	0.019	Planctomycetaceae	SAS	0.79	0.005
Burkholderiaceae	RBL	0.73	0.011	Actinobacteria	KWK+SAS	0.72	0.008
<i>Arcicella</i>	RBL	0.71	0.011	<i>Citrobacter</i>	KWK+SAS	0.59	0.014
<i>Legionella</i>	RBL	0.71	0.018	Chlamydiales	KWK+SAS	0.56	0.017
Acetobacteraceae	RBL	0.69	0.019	Unknown Proteobacteria	KWK+SAS	0.51	0.030
<i>Legionella</i>	RBL	0.69	0.019				
<i>Derxia</i>	RBL	0.69	0.019				

Table 3: Results of co-occurrence analyses for the dominant OTUs (20 reads, 3 sites). Significant results are presented in bold. SES refers to standardized effect size.

Geographic location	C-score	P	SES
SAS	0.37	<0.0001	13.66
KWK	1.54	<0.0001	8.70
BGR	0.45	0.39	0.84
NAS	1.04	<0.0001	8.19
RBL	0.36	0.015	2.87
REGION	35.7	<0.0001	25.4

Table 4: Topology of the thermokarst systems co-occurrence networks. Regional corresponds to a network built around the selected 294 OTUs whereas Hubs refers to a network where the most connected 24 OTUs from the whole network (SI Fig. 5A) were removed prior to this analysis. Grey shading refers to topology characteristics of Erdős–Rényi random networks of similar size.

Network parameter	Regional	Hubs
Nodes	248	224
Nodes random	248	224
Edges	968	433
Edges random	968	433
N. components	3	26
N. components random	1	4
Diameter (radius)	7 (1)	9 (1)
Diameter (radius) random	5 (4)	9 (6)
Degree	7.81	3.87
Degree random	7.81	3.93
Density	0.03	0.02
Density random	0.03	0.02
Heterogeneity	1.06	0.96
Heterogeneity random	0.34	0.48
Centralization	0.16	0.08
Centralization random	0.02	0.03
Clustering coefficient (Cl)	0.25	0.15
Clustering coefficient random (Clr)	0.03	0.02
Characteristic path length (L)	3.06	3.90
Characteristic path length random (Lr)	2.89	4.13
Log response ratio Cl	0.92	0.87
Log response ratio L	0.02	-0.02

Figure captions

Figure 1: (A) UPGMA clustering based on weighted and normalized UniFrac distances among bacterial community samples. Clustering statistics were computed using 100 jackknife replicates. (B) Principal coordinate analysis (PCoA) using UniFrac weighted distance metric. The SAS and KWK valleys are located in sporadic (highly degraded) permafrost landscapes, while the NAS and BGR valleys are located in discontinuous (less degraded) permafrost landscapes, and reference rock-based lakes are located in the RBL valleys. (C) Differences in the phylogenetic structure (NRI, net relatedness index) of bacterial communities among the different valleys. The solid black horizontal and vertical lines represent the mean and SD respectively. The dashed line represents the mean NRI value of NAS valley, with the 2 outliers excluded. Black dots represent individual pond and lakes.

Figure 2: Heatmap representation of habitat preference of the 30 most abundant (panel A) and most connected (panel B) bacterial OTUs. Habitat preference was determined by point biserial correlation. The connectivity of OTUs was defined by the degree (number of edges) of nodes from an association network (SI Fig. 5A). Taxonomic assignment of OTUs is provided at the lowest level of classification possible after BLASTn search in GenBank database.

Figure 3: Subnetworks organized around DOC (A) and phototrophic picoeukaryotes (B). Sub-networks were extracted from the entire co-occurrence network (SI Figure 5). In panel A, edge color refers to the type of relationship with significant connection between OTUs and both biotic and abiotic variables presented in black whereas relationships between bacterial taxa are presented in grey. In panel B, edge color is proportional to the association strength, with strong associations shown in black. The size of the nodes is proportional to node degree (the number of connections that a node has with other nodes).

Figure 4: Subnetworks organized around bacterial OTUs closely related to Chitinophagaceae. Panel A corresponds to the ensemble of co-occurrences between

members of Chitinophagaceae and other bacteria. Panel B refers the specific linkages between Chitinophagaceae and Actinobacteria. The size of the nodes is proportional to node degree (the number of connection that a node has with other nodes).

Figure 5: Associations between bacterial OTUs in permafrost thaw pond and lakes. (A) Co-occurrence between two representatives of Gammaproteobacteria that is partially explained by total nitrogen. (B) Non co-existence that is explained by the valley identity. (C) Co-occurrence between OTUs closely related to Actinobacteria ACK-M1 and Bacteroidetes Chitinophagaceae bacteria.

Figure 1

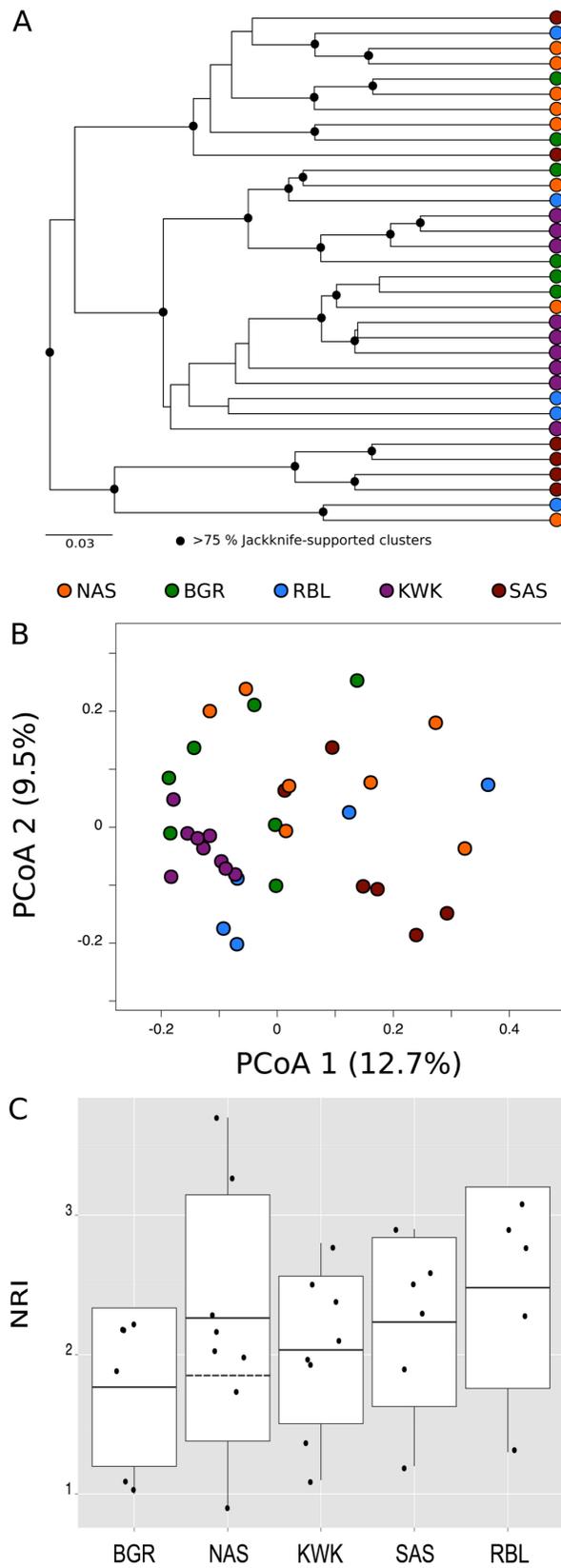


Figure 2

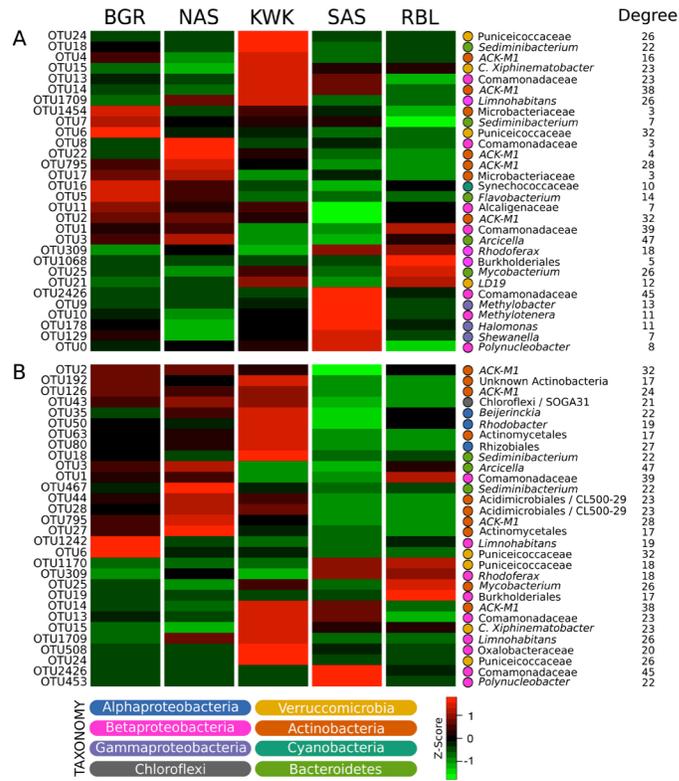


Figure 3

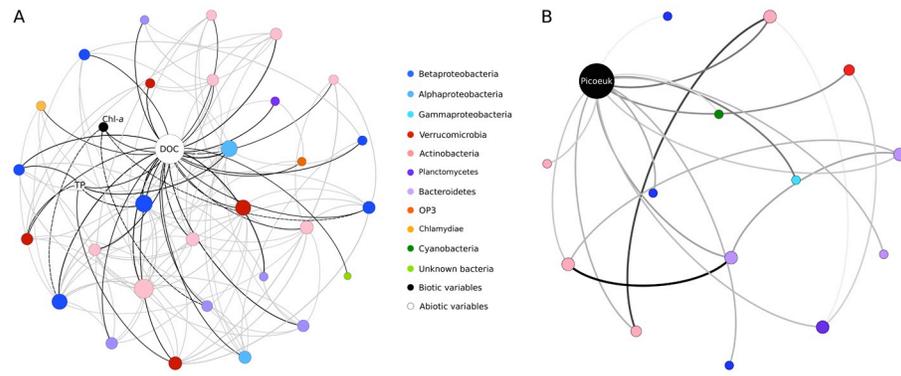


Figure 4

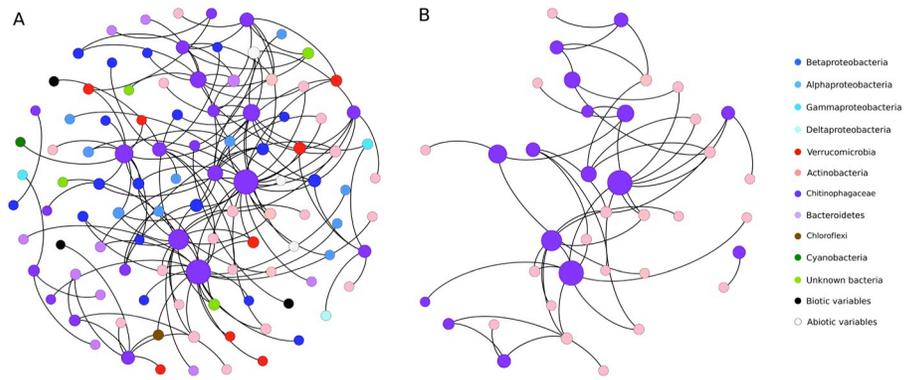


Figure 5

