

Running title: Extreme Metagenomics

1 In-field metagenome and 16S rRNA gene amplicon nanopore sequencing
2 robustly characterize glacier microbiota

3
4 Arwyn Edwards^{1,2*}, Aliyah R. Debbonaire^{1,2}, Samuel M. Nicholls^{1,2,3} Sara M.E. Rassner^{2,4},
5 Birgit Sattler⁵, Joseph M. Cook⁶, Tom Davy^{1,2}, André Soares^{2,4}, Luis A.J. Mur¹, Andrew J.
6 Hodson⁷

7
8 ¹Institute of Biological, Environmental & Rural Sciences (IBERS), Aberystwyth University,
9 Aberystwyth, SY23 3DD, UK; ²Interdisciplinary Centre for Environmental Microbiology,
10 Aberystwyth University, SY23 3DD, UK; ³Department of Computer Science, Aberystwyth
11 University, Aberystwyth, SY23 3DB, UK ⁴ Department of Geography & Earth Sciences,
12 Aberystwyth University, Aberystwyth, SY23 3DB, UK; ⁵Institute of Ecology, University of
13 Innsbruck, Innsbruck, Austria; ⁶Department of Geography, University of Sheffield, Sheffield,
14 UK. ⁷Department of Arctic Geology, University Centre in Svalbard (UNIS), Longyearbyen,
15 Svalbard.

16 *CORRESPONDENCE: Arwyn Edwards, Institute of Biological, Rural and Environmental
17 Sciences, Aberystwyth University, Aberystwyth, UK aye@aber.ac.uk

18
19 KEYWORDS: Metagenomics, Microbiome, Nanopore, MinION, Cryoconite, Glacier

20

21

Running title: Extreme Metagenomics

22 **ABSTRACT**

23 In the field of observation, chance favours only the prepared mind (Pasteur). Impressive
24 developments in genomics have led microbiology to its third “Golden Age”. However,
25 conventional metagenomics strategies necessitate retrograde transfer of samples from
26 extreme or remote environments for later analysis, rendering the powerful insights gained
27 retrospective in nature, striking a contrast with Pasteur’s dictum. Here we implement highly
28 portable USB-based nanopore DNA sequencing platforms coupled with field-adapted
29 environmental DNA extraction, rapid sequence library generation and off-line analyses of
30 shotgun metagenome and 16S ribosomal RNA gene amplicon profiles to characterize
31 microbiota dwelling within cryoconite holes upon Svalbard glaciers, the Greenland Ice Sheet
32 and the Austrian Alps. We show in-field nanopore sequencing of metagenomes captures
33 taxonomic composition of supraglacial microbiota, while 16S rRNA. Furthermore,
34 comparison of nanopore data with prior 16S rRNA gene V1-V3 pyrosequencing from the
35 same samples, demonstrates strong correlations between profiles obtained from nanopore
36 sequencing and laboratory based sequencing approaches. gene amplicon sequencing resolves
37 bacterial community responses to habitat changes. Finally, we demonstrate the fidelity and
38 sensitivity of in-field sequencing by analysis of mock communities using field protocols.
39 Ultimately, in-field sequencing potentiated by nanopore devices raises the prospect of
40 enhanced agility in exploring Earth’s most remote microbiomes.

Running title: Extreme Metagenomics

41 INTRODUCTION

42 Microbes drive biogeochemical processes at all scales within the biosphere(Falkowski et al.,
43 2008). The ubiquity of microbes within every manner of niche on Earth is underpinned by the
44 expansive range of microbial genomic diversity, spanning all three domains(Hug et al.,
45 2016). However, our view of microbial diversity on Earth is constantly changing as
46 technological advances reveal novel groups and associations of microbes(Rinke et al., 2013;
47 Hug et al., 2016)). Within recent decades, sequencing of DNA extracted from
48 environmental matrices, either amplified(Schmidt et al., 1991) or sequenced directly via
49 shotgun metagenomics(Tyson et al., 2004) makes the prospect of a predictive, multi-
50 scaled understanding of microbial interactions with the Earth system increasingly
51 tangible(Widder et al., 2016).

52 Typically, such investigations require the transfer of collected microbiota for nucleic acid
53 extraction, sequencing and bioinformatic analysis at facilities with a high level of
54 infrastructure. These strategies are well established. However, the necessity to transfer
55 samples from remote locations to the laboratory incurs several disadvantages, including the
56 potential loss or corruption of unique samples in transit as well as biases incurred by
57 taphonomic degradation during storage and subsequent extraction(Klein, 2015). In
58 particular, the delays incurred via this strategy means that the process of gaining
59 genomic insights to microbial processes within the natural environment is divorced
60 from the environment in which they occur, rendering the exploration of microbial
61 diversity a reactive, rather than proactive activity.

62 Pocket sized, USB-driven, nanopore DNA sequencers (e.g. the Oxford Nanopore
63 Technologies Ltd MinION) potentially offer a novel direction for the versatile
64 characterization of microbiomes. Highly portable DNA sequencing strategies promise the

Running title: Extreme Metagenomics

65 generation and analysis of DNA sequences within field settings, thus removing the risk of
66 sample degradation or loss in cold chain to a home laboratory and significantly accelerating
67 the process of generating insights.

68 While MinION sequencing has been applied for in-field genomic epidemiology studies of the
69 West Africa Ebola virus outbreak(Quick et al., 2016) and genome sequencing on the
70 International Space Station(Castro-Wallace et al., 2017) its application to in-field microbial
71 ecology applications has been limited. However, the potential for characterizing microbial
72 communities via nanopore-based shotgun metagenomics or amplicon sequencing on the
73 MinION platform has been established within laboratory-based sequencing experiments
74 using contrived or well-known communities(Brown et al., 2017; Kerkhof et al., 2017).
75 Astrobiologists have long used polar field sites as low-biomass environment analogues as a
76 testbed for life detection strategies, and the incorporation of MinION in trials for life
77 detection strategies offers some insight to the performance of MinION in remote and
78 challenging environments. This includes a trial of the MinION platform's technical
79 performance and endurance in Antarctica (Johnson et al., 2017) and the incorporation of
80 MinION-based metagenomics along with culturing as a proof-of-principle life detection
81 system in the Canadian High Arctic(Goordial et al., 2017). This latter study demonstrated the
82 potential utility of MinION sequencing in field settings but due to poor weather and limited
83 internet connectivity was limited to only one successful experiment where data could be
84 processed using a cloud-based bioinformatics platform.

85 Therefore, applications of nanopore sequencing in the field to date have typically required
86 labour-intensive sequencing library preparation following protocols developed for laboratory
87 use, and data analysis required base-calling and/or annotation using remote servers. Reliance
88 on cold chain transfer of reagents and sequencing flow cells also presents an important
89 limitation. However, the implementation of shotgun metagenomics and 16S rRNA gene

Running title: Extreme Metagenomics

90 amplicon analysis for in-field sequencing of complex microbiomes in conditions typical for
91 microbial ecologists in remote environments requires these obstacles to be overcome.

92 In this study we show the applicability of nanopore sequencing for rapid taxonomic
93 characterization and comparison of microbial communities in remote environments.
94 Moreover, we benchmark the performance of in-field sequencing and bioinformatics
95 approaches used by sequencing of mock communities analysed following in-field protocols
96 and re-analysis of samples previously investigated by second-generation sequencing. For our
97 analyses we selected cryoconite, a photo- and heterotrophic microbe-mineral aggregate
98 darkening the ice surfaces of glaciers and ice sheets (Cook et al., 2015) as target microbiota.
99 Cryoconite aggregates are home to a diverse range of microbial life, with estimated diversity
100 and activity rates comparable to certain soils (Anesio et al., 2009; Cameron et al., 2012) in
101 spite of its icy environs. We show shotgun metagenomic and 16S rRNA gene amplicon
102 sequencing using rapid library preparation protocols and fast, laptop-based taxonomic
103 classification tools are capable of robustly characterizing and comparing microbial
104 communities while in the field. Our results raise the prospect of investigating Earth's
105 microbiomes at source.

106

107

Running title: Extreme Metagenomics

108 **MATERIALS AND METHODS**

109 Sampling

110 Sampling and sequencing was performed in Svalbard and Greenland, while archived
111 materials from three Svalbard glaciers and two alpine glaciers were used in the benchmarking
112 experiments. In all cases, cryoconite samples were collected using a disinfected turkey baster
113 which was used for the aspiration of debris to 15 mL sterile centrifuge tubes and transferred
114 within hours on ice in ambient temperatures of +2-4°C to the field lab or camp where they
115 were either extracted directly or stored in a -20°C freezer. While a common DNA extraction
116 protocol is employed, over the course of this study, protocols and tools for nanopore
117 sequencing have advanced rapidly and we describe the implementation of nanopore
118 sequencing in remote locations with differing levels of infrastructure. Each experiment is
119 therefore described in turn and summarized in Table 1.

120 In-field DNA extraction

121 For cryoconite studies in the present study, community genomic DNA was extracted directly
122 from cryoconite debris in field conditions. PowerSoil® DNA Isolation kits (MoBio, Inc.) are
123 commonly used to extract high quality DNA from environmental matrices rich in nucleic acid
124 processing enzyme inhibitors, including glacial samples (Edwards et al., 2011; Cameron et
125 al., 2012; Musilova et al., 2015; Stibal et al., 2015; Lutz et al., 2016). We therefore opted to
126 amend the PowerSoil® DNA isolation protocol for use within a minimalistic field laboratory
127 to generate DNA extracts compatible with nanopore sequencing protocols.

128 Generally, the protocol was implemented as per the manufacturer's instructions with the
129 following variations: The starting material was increased from 0.25 g to 0.27-0.29 g wet
130 weight to account for moisture content prior to bead-beating within PowerBead tubes. Bead-

Running title: Extreme Metagenomics

131 beating was conducted using a Vortex Genie 2 (Scientific Industries, Inc) fitted with a MoBio
132 (MoBio, Inc.) tube adapter or a IKA Works, Inc MS2 S8 Minishaker. Since the PowerSoil®
133 protocol typically uses benchtop centrifuges capable of 15-17,000 ×g, the use of a Gilson, Ltd
134 GmCLab Microcentrifuge capable of generating <2,900 ×g required elongated spin times of
135 four minutes for protein precipitation following the addition of solution C2 and inhibitor
136 removal following the addition of solution C3. For all other centrifugation steps, spin times of
137 one minute were sufficient. To increase DNA concentration, DNA was eluted in 70 µL
138 solution C6, after a five minute incubation at room temperature. Two microlitre aliquots of
139 each extract were than immediately quantified using Qubit 2.0 Fluorimeter dsDNA High
140 Sensitivity assays (Invitrogen, Ltd) prior to sequencing. In the case of Greenland
141 metagenomics, fluorimetry was performed with the Qubit device powered using a portable
142 power-pack (PowerAdd, Inc.)

143 Shotgun metagenomics

144 Svalbard: shotgun metagenomics experiment, Foxfonna ice cap.

145 On 18 August 2016, cryoconite debris from four cryoconite holes was collected as described
146 above on the un-named outlet glacier of Foxfonna ice cap (78° 08'N, 16° 07' E; (Rutter et al.,
147 2011; Gokul et al., 2016).) Cryoconite was thawed the following day and DNA extracted as
148 described. The analyses were performed in a room without prior laboratory infrastructure.
149 Equimolar pools of cryoconite DNA provided 185 ng DNA used for a transposase-based 1D
150 rapid sequencing library preparation exactly as specified by Oxford Nanopore Technologies,
151 Ltd (Rapid Sequencing of genomic DNA for the minION device using SQK-RAD001) using
152 a PCR cycler for the 30°C and 75°C incubation steps in a Hybaid Omni-E PCR cycler.

153 1D sequencing was directly performed using MinION Mk 1 device containing a R9 flow cell
154 (Oxford Nanopore Technologies, Ltd) loaded conventionally in 150 µL sequencing mix. The

Running title: Extreme Metagenomics

155 MinION Mk 1 device was controlled using MinKNOW 1.0.5 (Oxford Nanopore
156 Technologies, Ltd). Sequencing and read processing were performed on a Lenovo®
157 ThinkPad® X1 Carbon running Windows 10 with Intel® Core™ i7-5600U Processor, 8GB
158 RAM and a 512GB solid state drive. 1D basecalling was performed remotely via the Oxford
159 Nanopore Technologies, Ltd. Metrichor platform. Template strand nanopore reads in .fast5
160 format were converted to fasta using the pore_parallel GUI of poRe 0.17 (Watson et al.,
161 2015) running in R 3.3.1.

162 Greenland Ice Sheet margin: shotgun metagenomics experiment, Russell glacier.

163 On 16 June 2017, cryoconite debris from four cryoconite holes was collected as described
164 from four cryoconite holes on Russell glacier at the south western margin of the Greenland
165 Ice Sheet (N67° 8'08.871' 05 ° 08'05.016' W(Edwards et al., 2014b)) DNA extraction and
166 sequencing was performed in a field camp tent with AC generator power for DNA extraction
167 only. Equimolar pools of cryoconite DNA provided 185 ng DNA used for transposase based
168 1D rapid sequencing library preparation as specified by Oxford Nanopore Technologies, Ltd
169 (Rapid Sequencing of genomic DNA for the minION device using SQK-RAD002) but
170 holding the library tube in gloved hands for the 30°C incubation step and immersion in hot
171 water stored in an insulated mug for 75°C inactivation.

172 The library (75 µL) was loaded in the SpotON port of a R9.4 flow cell (Oxford Nanopore
173 Technologies, Ltd) run using a MinION MK1 device. Due to the cold (<0°C) ambient
174 temperatures, to maintain optimal sequencing temperature, the MinION device was enclosed
175 in an insulated cool bag during sequencing (Tesco, Ltd). Sequencing and read processing
176 were performed on a Dell XPS15 laptop running Windows 10 Intel® Core™ i7-6700HQ
177 Processor, 32GB RAM and a 1TB solid state drive, running on internal battery. Sequencing
178 was controlled using MinKNOW 1.6.11 (modified for offline use, courtesy Oxford Nanopore

Running title: Extreme Metagenomics

179 Technologies, Ltd) and 1D basecalling performed locally with Albacore v1.1.0 (Oxford
180 Nanopore Technologies, Ltd).

181 Benchmarking: shotgun metagenomics experiment, Rotmoosferner, Austrian Alps.

182 We sought to compare the performance of in-field nanopore metagenome sequencing with
183 established strategies for metagenome sequencing and compare the performance of rapid
184 library kits based upon aqueous reagents used in the field study with the current state of the
185 art, namely rapid library kits based upon freeze-dried reagents.

186 Therefore, we conducted nanopore sequencing of samples with corresponding, publicly
187 available metagenome data. Samples RF1 and RF6, from alpine cryoconite were collected
188 from the surface of Rotmoosferner glacier in the Austrian Alps in September 2010. A
189 shotgun metagenomic sequencing library of the same two samples in an equimolar pool with
190 12 other cryoconite community genomic DNA samples from Rotmoosferner was analysed by
191 Illumina sequencing previously (Edwards et al., 2013b) and is available at MG-RAST
192 (4 491 734.3).

193 Community genomic DNA was extracted in 2017 from 250 mg cryoconite sediment from
194 samples RF1 and RF6 stored at -80°C using the field-modified PowerSoil method in our
195 home laboratory. To generate rapid 1D libraries, 400 ng DNA from the samples were used in
196 transposase-based protocol as specified by Oxford Nanopore Technologies, Ltd (Rapid
197 Sequencing of genomic DNA for the minION device using SQK-RAD003) with incubation
198 steps in a PCR Cycler (G-Storm Direct, Ltd). Library was loaded and sequenced on a R9.4
199 flow cell as described above for the *Greenland* experiment, using MinKNOW 1.7 and
200 Albacore v2.02 for sequencing and basecalling respectively.

Running title: Extreme Metagenomics

201 Storage and transportation of molecular biology reagents and sequencing flow cells at cold
202 temperatures represents a critical limitation to the utility of nanopore sequencing for
203 characterizing microbial communities in remote locations. We therefore tested lyophilized
204 versions of the rapid library kit (SQK-LRK001) and ambient-shipped R9.4 flow cells
205 provided by Oxford Nanopore Technologies Ltd. in a repetition of the above experiment. The
206 experiment was performed without laboratory resources. Briefly, 400 ng of community DNA
207 in a 10 μ L volume was used to suspend lyophilized fragmentation mix, with incubation at
208 room temperature with inactivation and immersion in hot water stored in an insulated mug for
209 80°C inactivation for one minute each. The resulting solution was incubated with rapid
210 adapter for five minutes before gentle mixing with 65 μ L priming buffer and flow cell
211 loading before sequencing as described above.

212 Benchmarking: sequencing of mock community DNA using lyophilized rapid kit

213 Finally, to establish the fidelity with which in-field DNA sequencing can capture microbial
214 diversity, we performed DNA sequencing on the ZymoBIOMICS™ Microbial Community
215 standard (provided as DNA, D3606, Zymo Research, Inc.) using the lyophilised rapid kit and
216 field protocols, albeit in a laboratory setting.

217 The ZymoBIOMICS™ Microbial Community standard comprises a mixture of eight bacterial
218 species (*Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus*
219 *fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Bacillus*
220 *subtilis*) each comprising 12% of the DNA by mass, and two yeast species (*Saccharomyces*
221 *cerevisiae* and *Cryptococcus neoformans*) each comprising 2% of the DNA by mass. The
222 manufacturers state that the mixture contains <0.01% foreign microbial DNA and the
223 composition is stable to within a relative abundance deviation of 15% on average. Further
224 details on the ZymoBIOMICS™ Microbial Community standard are available at:

Running title: Extreme Metagenomics

225 https://www.zymoresearch.eu/media/amasty/amfile/attach/_D6305_D6306_ZymoBIOMICS_
226 [Microbial_Community_DNA_Standard_v1.1.3.pdf](#)

227 Briefly, 400 ng of the ZymoBIOMICS™ Microbial Community standard in a total volume of
228 10 µL was used, exactly as described above for lyophilised rapid kit sequencing.

229 Bioinformatics for shotgun metagenomics

230 For all experiments, sequencing reads were analysed using Kaiju(Menzel et al., 2016), a
231 taxonomic classifier which is offered as a webserver (<http://kaiju.binf.ku.dk>.) or as a
232 standalone version. Kaiju seeks protein-level matches in all possible reading frames using the
233 Burrows-Wheeler transform and has been demonstrated to be rapid yet sensitive (Menzel et
234 al., 2016). We report Kaiju analyses conducted using the NCBI *nr+euk* database (version: 05
235 May 2017) containing 103 million protein sequences from microbes, viruses and selected
236 microbial eukaryotes with low complexity filtering and in *greedy* mode (Minimum match:
237 11; minimum match score: 75; allowed mismatches: 5). Nanopore sequencing data from in-
238 field metagenomes are available at the European Nucleotide Archive (ENA) accessions
239 ERR2264275 - ERR2264278 while mock community metagenomes are available at ENA
240 PRJEB30868.

241 16S rRNA gene sequencing

242 Svalbard: 16S ribosomal RNA gene sequencing, Vestre Brøggerbreen

243 On 10 July 2017, cryoconite debris was collected from six cryoconite holes on Vestre
244 Brøggerbreen (78° 54'7 N, 011°43'8 E) on Svalbard. Three cryoconite holes (VB1-3) were
245 covered by snow and superimposed ice while three cryoconite holes (VB4-6) were exposed.
246 To test the hypothesis that the presence or absence of snow/ice cover incurs changes in the
247 bacterial community structure, 16S ribosomal RNA gene sequencing was conducted. Samples

Running title: Extreme Metagenomics

248 were transferred to the NERC Arctic Research Station Ny Ålesund within three hours and
249 DNA extracted and quantified as described above within the field lab.

250 Bacterial 16S rRNA genes were amplified from 50 ng DNA per sample diluted to 10 µL
251 nuclease free water in 50 µL with 1 × LongAmp Taq master mix (New England Biolabs,
252 Inc), 1 µL 16S rRNA gene barcoded primer (ONT-SQK-RAB-201, Oxford Nanopore
253 Technologies, Ltd) and 14 µL nuclease free water. Each sample was allocated to a barcoded
254 primer. A no template control (nuclease free water) and extract control (blank extract) were
255 amplified and sequenced in parallel. PCR was conducted as specified in the Oxford Nanopore
256 Technologies protocol for rapid, barcoded sequencing of 16S rRNA genes (ONT-SQK-RAB-
257 201, Oxford Nanopore Technologies, Ltd) with the modification that 30 cycles of PCR were
258 performed. PCR was performed using a 8 well MiniPCR device (Amplify, Inc.) controlled
259 by a Windows laptop. The resulting PCR products were purified using magnetic beads
260 (Ampure XP, Beckman Coulter, Ltd) and eluted in 10 µL buffer (10 mM Tris, 50 mM NaCl,
261 pH 8.0) prior to Qubit quantification. Equimolar quantities of each barcode were pooled and
262 sequenced using a R9.4 flow cell as described above for the *Greenland* experiment prior to
263 1D basecalling and barcode demultiplexing in Albacore v1.10.

264 Benchmarking: nanopore 16S rRNA gene discrimination of Arctic and alpine cryoconite

265 To compare the performance of in-field nanopore 16S rRNA gene sequencing with
266 established strategies for 16S rRNA gene profiling, we conducted nanopore sequencing of
267 cryoconite samples with corresponding, publicly available 16S rRNA gene data spanning
268 Svalbard and Austrian glaciers. V1-V3 16S rRNA amplicon pyrosequencing data for these
269 samples are available at available at EBI-SRA (PRJEB5067-ERP004426). Details of sample
270 collection, pyrosequencing and data processing are described elsewhere (Edwards et al.,
271 2014b).

Running title: Extreme Metagenomics

272 Ten samples (Svalbard: three pairs of samples from Austre Brøggerbreen [AB], Midtre
273 Lovénbreen [ML] and Vestre Brøggerbreen [VB]. Austria; two pairs of samples from
274 Gaisbergferner[GB] and Rotmoosferner[RM]) archived at -80°C since collection were used
275 to generate fresh DNA extracts as described above. Bacterial 16S rRNA genes were
276 amplified and sequenced exactly as described for the Svalbard 16S ribosomal RNA gene
277 sequencing experiment, with the exception that a 96 well PCR cycler (G-Storm Direct, Ltd)
278 was used for amplification and Albacore v2.02 used for basecalling and demultiplexing by
279 barcode. Samples were assigned an individual barcode primer, with co-amplification and
280 sequencing of no-template and extract controls.

281 Benchmarking: 16S rRNA gene nanopore sequencing of mock communities.

282 To evaluate the accuracy and sensitivity of 16S rRNA gene nanopore sequencing we
283 analysed mock communities. The ZymoBIOMICS™ Microbial Community Standard was
284 used to represent a baseline community of eight members with 16S rRNA genes. To
285 establish the accuracy of taxonomic assignment using uncorrected 16S rRNA gene reads and
286 the sensitivity of our protocol to be able to detect taxa present above and below 1%
287 theoretical relative abundance, we spiked the ZymoBIOMICS™ Microbial Community
288 Standard with genomic DNA of *Micrococcus luteus* NCTC2665, an actinobacterial strain
289 with a 2.65Mbp genome and only one functional 16S rRNA gene (Young et al. 2010). *M.*
290 *luteus* DNA was extracted and quantified as above, from an overnight nutrient broth culture
291 stored on ice for one hour prior to extraction. Genomic DNA from *M. luteus* was spiked in to
292 create mixtures where the theoretical relative abundance of the *M. luteus* 16S rRNA gene was
293 0%, 0.4%, 0.8% and 1.2% of the total bacterial 16S rRNA genes. Otherwise, amplification
294 and sequencing of 16S rRNA genes were conducted exactly as described above for the
295 comparison of Arctic and alpine cryoconite with the exception that negative and extraction
296 controls were not sequenced. It should be noted that two members of the community,

Running title: Extreme Metagenomics

297 *Escherichia coli* and *Salmonella enterica* share highly conserved (typically well above 97%
298 identity, e.g. Fukushima et al. 2002) 16S rRNA genes therefore these taxa were treated as a
299 single entity in data processing.

300 16S rRNA gene data analysis

301 Reads were base-called and de-multiplexed using Albacore v1.10 and v2.02 for
302 implementation and benchmarking experiments, which were converted to fasta for taxonomy
303 assignment. In developing a strategy for nanopore 16S rRNA amplicon analysis, our primary
304 considerations were to leverage the short turnaround time and in-field flexibility of the
305 MinION platform while mitigating the impacts of relatively high error rate per base. We
306 therefore opted to directly assign higher-level taxonomy to Albacore demultiplexed reads via
307 the SINTAX taxonomic classifier in usearch v10.0.240(Edgar, 2016) against taxa in a
308 species-level identity curated database of Ribosomal Database Project version 16 taxa. A
309 confidence level of 0.75 was chosen for each taxonomic assignment. The taxonomy
310 assignment is available as supplementary data.

311 The numbers of reads assigned per taxon were counted in MS Excel and the relative
312 abundance of reads per taxon used for separate downstream analysis. Multivariate analyses of
313 bacterial community structure using Bray-Curtis distances of fourth root transformed taxon
314 relative abundance data were performed in Primer-6.1.12 & Permanova+1.0.2 (Primer-E, Ltd,
315 Ivybridge, UK). Principal Coordinates Analysis (PCoA) and group-average Hierarchical
316 Cluster Analysis (HCA) were used as unsupervised data exploration, while hypotheses were
317 tested using PERMANOVA with unrestricted permutation of raw data and 9,999 ordinations.
318 Reads from the implementation, comparison and mock community experiments are available
319 from ENA at ERR2264279 - ERR2264283, ERR2264284 - ERR2264293 and PRJEB30868.

320

Running title: Extreme Metagenomics

321 **RESULTS AND DISCUSSION**

322 We report the successful implementation of nanopore sequencing for generating taxonomic
323 profiles of microbiota in field environments, both from shallow-depth shotgun metagenome
324 and targeted, 16S rRNA gene amplicon sequencing approaches. Our analyses reveal
325 nanopore sequencing generates taxonomic profiles coherent with conventional workflows in
326 molecular microbial ecology. These results raise the prospect that highly portable DNA
327 sequencing can be applied to characterise microbiomes rapidly while researchers are
328 deployed in remote field sites and thus help inform experimental or survey planning and
329 analysis within the field.

330 In field metagenomics of Arctic cryoconite

331 Here, we report the generation of two metagenomes from Svalbard and Greenland cryoconite
332 by in-field nanopore sequencing, and the taxonomic classification of microbial diversity (**FIG**
333 **1**).

334 For the *Svalbard* metagenomics experiment, community DNA was extracted from Foxfonna
335 cryoconite and sequenced using rapid library preparation, with 3,514 reads successfully
336 basecalled. Of these, through analysis using the Kaiju classifier, 2,305 reads could be
337 assigned to named taxa, of which 2,265 reads were assigned as bacterial in origin. Consistent
338 with earlier V1-V3 region 16S rRNA gene amplicon sequencing of cryoconite from Foxfonna
339 ice cap (Gokul et al., 2016) the community was dominated by *Proteobacteria*
340 (*Alphaproteobacteria* and *Betaproteobacteria*) followed by *Actinobacteria*, *Cyanobacteria*
341 and *Bacteroidetes* (**FIG 1**). Importantly, the most abundant species-level match was to
342 *Phormidesmis priestleyi*, representing 3% of the reads assigned to any taxon. The
343 cyanobacterium *Phormidesmis priestleyi* is frequently detected in Arctic cryoconite and is
344 thought to act as an ecosystem engineer of cryoconite granules, with its cyanobacterial

Running title: Extreme Metagenomics

345 filaments entangling inorganic debris to form darkened microbe-mineral aggregates on the
346 ice surface (Langford et al., 2010; Cook et al., 2016; Gokul et al., 2016; Uetake et al., 2016).

347 The *Svalbard* metagenomics experiment represents an initial use of rapid library preparations
348 to generate taxonomic profiles in field environments, with base-calling and subsequent
349 analysis performed using online resources. Additionally, this experiment represents sample-
350 to-preprint communication of its initial analyses within 23 days of sample collection
351 (Edwards et al., 2016). Considering the doubling times of Svalbard cryoconite sedimentary
352 bacterial communities are in the same order (Anesio et al., 2010) this represents the
353 metagenomic characterization of a microbial community within its generation time.

354 The reliance of our *Svalbard* metagenomics experiment described above on internet access
355 and mains electricity represents an important limitation. We therefore sought to conduct a
356 second experiment based in a typical field camp setting. Our *Greenland* metagenome
357 sequencing experiment entailed collection of cryoconite from the margin of the Greenland Ice
358 Sheet and its sequencing. DNA extraction was performed with access to a portable generator,
359 but generator failure required all subsequent steps to be performed on battery power.
360 Quantification powered by a portable power-pack while library preparation was performed
361 without power, and DNA sequencing was performed on laptop batteries. Freezing air
362 temperatures presented challenges for maintaining liquid reagents and optimal MinION
363 temperatures, resolved by storing reagents warmed by body warmth in a down jacket, and
364 operating the MinION in an insulated shield. These limitations constrained the endurance of
365 the sequencing experiment, but of 2,372 reads sequenced, 796 reads were assigned to
366 taxonomy and 692 reads matched bacterial taxa (FIG 1). The community revealed was
367 dominated by *Cyanobacteria*, *Proteobacteria* (*Alphaproteobacteria*, *Betaproteobacteria*,
368 *Gammaproteobacteria*) and *Actinobacteria*, again consistent with sequence data from
369 Greenlandic cryoconite(Edwards et al., 2014b; Musilova et al., 2015; Cook et al., 2016). The

Running title: Extreme Metagenomics

370 most abundant species within the profile, *Phormidesmis priestleyi*, represented 6% of the
371 reads assigned to any taxon. Our *Greenland* metagenome experiment therefore represents the
372 in-field metagenomic sequencing of a microbial community in the resource-limited settings
373 typical of Arctic field camps.

374 Benchmarking in-field metagenomics: the utility of freeze-dried reagents for sequencing

375 An important limitation to all above experiments is the necessity of maintaining a cold-chain
376 for sequencing reagents and nanopore flow cells. This presents logistical challenges for
377 characterizing microbial communities in highly remote locations. We therefore tested an
378 early-release lyophilized field kit provided by Oxford Nanopore Technologies, Ltd, now
379 available as SQK-LRK001. We compared its performance to cold-chain dependent aqueous-
380 based reagents (SQK-RAD003) for rapid generation of sequencing libraries. To do so, we re-
381 sequenced cryoconite DNA samples from Rotmoosferner, an alpine glacier, which had
382 previously been characterized using Illumina sequencing (Edwards et al., 2013a). Two
383 experiments were performed. The first experiment was performed using field protocols
384 implemented in a laboratory setting while the second experiment used reagents and R9.4 flow
385 cells adapted for ambient shipping and storage and stored at room temperature for at least
386 five days. Kaiju classified 18,694 out of 31,538 reads generated using the lyophilised reagent
387 kit, of which 18008 were assigned to the bacterial domain (FIG 1) while 2847 out of 6844
388 reads generated by the aqueous kit were classified. Comparison of the aqueous (Austria-
389 RAD003) and lyophilised (LRK001) libraries (FIG 1) from the same samples showed highly
390 coherent taxonomic profiles at higher grade taxon (Pearson $r=0.97$, $p<0.001$) and species
391 levels (FIG1A , FIG 1B). The Kaiju-profiled communities were dominated by *Proteobacteria*
392 (*Betaproteobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*) with *Burkholderiales*
393 prominent within the community, consistent with previously reported Illumina-based shotgun
394 metagenomics analyses of the same samples (Edwards et al., 2013a). *Cyanobacteria* were

Running title: Extreme Metagenomics

395 dominated by *Chamaesiphon minutus* matching reads at 7% of all reads sequenced in both
396 samples. *Chamaesiphon* sp. are a prevalent cyanobacterial species within mountain glacier
397 cryoconite (Segawa et al. 2017) therefore the detection of *C. minutus* as the dominant taxon
398 within the alpine metagenomic dataset is likely valid.

399 Benchmarking in-field metagenomics: sequencing of mock communities

400 To provide an estimation of the fidelity and reproducibility of in-field shotgun metagenomics
401 we sequenced the ZymoBIOMICS™ Microbial Community Standard using the lyophilised
402 field kit (LRK001) in three sequencing experiments (FIG 2). For the three experiments 94%
403 of 29,399 reads, 95% of 52,000 reads and 94% of 29,084 reads were respectively classified
404 by Kaiju. The notably higher percentage of reads classified compared to experiments with
405 environmental communities likely represents the stronger representation of the well-known
406 taxa in the mock community within sequence databases compared to taxa from poorly
407 described environments. All the expected species from the mock community were detected in
408 the sequencing runs at family level. Within the *Enterobacteriaceae* reads were assigned to the
409 genera *Escherichia* and *Salmonella* which could be expected within the community, but also
410 the closely-related genus *Shigella* at a relative abundance of 1%. The taxonomic composition
411 of the communities was highly consistent between sequencing experiments (FIG 2A), and
412 correlated well with the expected relative abundances (FIG 2B, Pearson $r=0.67$, $p=0.02$).

413 The performance of Kaiju with nanopore data in our experiments matches the expectations of
414 performance for Illumina data as described by its authors (Menzel et al., 2016). Since Kaiju
415 uses protein-level matches to all possible reading frames it likely represents an effective tool
416 for the rapid characterization of the taxonomic composition and implicit functional potential
417 of microbial communities analysed by rapid library in-field nanopore sequencing of
418 metagenomic DNA.

Running title: Extreme Metagenomics

419 The performance of lyophilized field kit and ambient flow cells used for sequencing natural
420 and mock communities using field protocols described above raises the prospect of
421 metagenomic characterization of microbial communities in highly resource limited, remote
422 environments.

423 In field 16S rRNA gene sequencing

424 Amplicon sequencing of 16S rRNA genes represents a backbone technique in microbial
425 ecology for the culture-independent investigation of microbial diversity (eg(Thompson et al.,
426 2017)). In contrast with short read second-generation high throughput sequencing techniques,
427 full-length 16S rRNA gene sequencing is possible with using nanopore(Li et al., 2016;
428 Kerkhof et al., 2017) in laboratory settings. Therefore, we anticipated MinION based in-field
429 sequencing of barcoded 16S rRNA genes could permit the semi-quantitative comparison of
430 bacterial communities while in the field.

431 To test the feasibility of using barcoded 16S rRNA gene MinION sequencing to compare
432 bacterial communities we analysed cryoconite holes which were either open to the
433 atmosphere (VB1-3; n=3 FIG3 B) or covered by a layer of snow and superimposed ice (VB4-
434 6; n=3 FIG3 B). On Arctic glaciers, cryoconite holes are seasonally open to the atmosphere in
435 summer, permitting photosynthesis as a dominant carbon fixation route and supporting a
436 diverse microbial community (Cook et al. 2015) Little is known of the dynamics of
437 cryoconite communities before the seasonal recession of snow cover exposes the cryoconite
438 hole, but it is likely the snow cover attenuates the flux of photosynthetically available
439 radiation. We sought to test the hypothesis that the presence or absence of snow/ice cover
440 incurs changes in the bacterial community structure.

441 We amplified and co-sequenced a blank extraction control and a blank PCR control to detect
442 the potential impacts of contamination arising from conducting 16S rRNA gene PCR (Salter

Running title: Extreme Metagenomics

443 et al., 2014), as preparing amplifications in a field laboratory setting may pose additional
444 risks of contamination. A limitation of our experiment was the number of samples which
445 could be processed in a single run, given the capacity of the eight-well PCR portable PCR
446 cyclers used. DNA was extracted and amplified from six samples, but PCR product from one
447 sample (VB4, open) was lost during bead clean-up, resulting in sequencing failure for that
448 sample.

449 Following basecalling and demultiplexing, 20,000 reads per sample were subjected to
450 taxonomic analysis. Only seven sequences were returned for each of the negative controls,
451 indicating the likely minimal impact of contamination in the field lab setting.

452 Although error-correcting approaches are in development for 16S rRNA gene nanopore
453 sequencing (Li et al., 2016; Calus et al., 2018), considering the goal of preliminary
454 characterization based upon rapid protocols compatible with field use, we opted to proceed
455 with uncorrected reads, directly assigning taxonomy to each read and treating the cumulative
456 abundance of reads matching discrete higher-level taxonomic affiliations as phylotypes.

457 Using the SINTAX classifier algorithm (Edgar, 2016) trained on a highly curated version of
458 the RDP database, reads were classified within ca. 10 minutes on a laptop computer without
459 the need for internet access. At a confidence level of 0.75, 15,643 reads per barcode were
460 assigned to bacterial taxa on average (range: 13,017-17,593, 1 SD = 1725 reads). All
461 community profiles were strongly dominated by *Cyanobacteria* at the phylum level with
462 prominent contributions from *Proteobacteria* (*Betaproteobacteria* and *Alphaproteobacteria*)
463 and *Bacteroidetes* broadly consistent with prior amplicon studies of Svalbard cryoconite (FIG
464 3). At both the phylum and family level, cryoconite communities ordinated and clustered
465 clearly according to the open or closed status of the cryoconite hole (principal coordinates
466 and hierarchical cluster analysis of fourth root transformed Bray-Curtis distances, FIG 4 A-

Running title: Extreme Metagenomics

467 **D**). At the family level, significant differences were apparent in the fourth root transformed
468 Bray Curtis distances of taxon relative abundance (PERMANOVA: $F=6.5$, $p(\text{MC})=0.02$).
469 Our analyses are consistent with an influence of snow/ice cover on total bacterial community
470 structure within cryoconite aggregates, however the increased abundance of cyanobacteria in
471 the closed cryoconite holes is intriguing; it may be that the lower light levels in holes with
472 thin snow/ice cover compared to open holes mitigate photoinhibitory effects (Cook et al.
473 2015) presenting a hypothesis which could be investigated further as a result of in-field
474 sequencing. Our results underscore the potential for in-field 16S rRNA gene sequencing to
475 inform investigations while deployed in remote environments.

476 Benchmarking in-field 16S rRNA gene sequencing

477 To establish whether in-field 16S rRNA gene sequencing implemented as above offers results
478 coherent with established, second-generation sequencing approaches we implemented our
479 nanopore protocol on DNA samples which had previously been characterized by V1-V3 16S
480 rRNA gene pyrosequencing (Edwards et al., 2014b). Our sample set comprised ten cryoconite
481 samples from three Svalbard glaciers ($n=2$ each) and two alpine glaciers ($n=2$ each). Negative
482 extraction and PCR controls as above produced one and four reads respectively. From 24,000
483 demultiplexed reads per sample, on average 20,882 reads were assigned to bacterial
484 taxonomy (range: 9,021-24,000, 1 SD =6,518, FIG 5). *Cyanobacteria* with prominent
485 contributions from *Proteobacteria* (*Betaproteobacteria* and *Alphaproteobacteria*) and
486 *Bacteroidetes* dominated the phylum level taxonomic distributions of the community. At the
487 phylum level, significant differences were observed between Arctic and alpine cryoconite
488 communities (PERMANOVA; $F=2.8866$, $p=0.02$) with discrete ordination apparent
489 (principal coordinates analysis of Bray-Curtis distances, FIG 5). At the improved taxonomic
490 resolution of family level assignments, the effect of parent glacier is highly significant
491 (PERMANOVA; $F=5.032$, $p=0.001$). In this regard, the outcomes of nanopore-based 16S

Running title: Extreme Metagenomics

492 rRNA gene sequencing are highly coherent with prior pyrosequencing analyses(Edwards et
493 al., 2014b).

494 To further compare the assignment of taxa following nanopore-based 16S rRNA gene
495 sequencing with the pyrosequencing dataset of the same samples, we correlated the log
496 relative abundance of dominant taxonomic groups (FIG 6). Significant, strongly positive
497 Pearson correlations between the log relative abundances of key taxonomic groups were
498 apparent between nanopore and pyrosequencing data: *Acidobacteria* ($r=0.82$, $p=0.004$),
499 *Alphaproteobacteria* ($r=0.81$, $p=0.004$), *Betaproteobacteria* ($r=0.90$, $p<0.001$), *Bacteroidetes*
500 ($r=0.70$, $p=0.02$), and *Firmicutes* ($r=0.90$, $p<0.001$). The ratios of
501 *Alphaproteobacteria:Betaproteobacteria*, previously identified as a discriminator between
502 Arctic and alpine cryoconite (Cook et al., 2015) were highly correlated between nanopore
503 and pyrosequencing datasets ($r=0.90$, $p<0.001$). In summary, nanopore sequencing of 16S
504 rRNA genes conducted according to in-field protocols correlates well with insights derived
505 from second-generation sequencing performed in laboratory settings.

506 Benchmarking: 16S rRNA gene nanopore sequencing of mock communities.

507 Our in-field 16S rRNA gene analyses are limited by two caveats. Firstly, to provide rapid, in-
508 field insights we must currently rely upon uncorrected nanopore 1D reads which are therefore
509 subject to a relatively high error rate. In this study we anticipated the availability of full-
510 length 16S rRNA gene reads and hence stronger phylogenetic signal relative to the random
511 error profile of nanopore-induced error, coupled with the aggregation of reads at higher
512 taxonomic levels would amortize the effect of diminished taxonomic resolution. Secondly,
513 the depth of sequencing may be relatively shallow compared to laboratory-based sequencing.

514 Therefore, we sought to evaluate the fidelity, reproducibility and sensitivity of in-field 16S
515 rRNA gene sequencing by using ZymoBIOMICS™ Microbial Community Standard mock

Running title: Extreme Metagenomics

516 communities where a single species, *Micrococcus luteus*, had been spiked at 16S rRNA gene
517 concentrations in the range 0-1.2% (FIG 7). A total of 220,051 reads were de-multiplexed,
518 with an average of 18,375 reads per sample (± 1 SD=3802, range 13,165-24,858). We
519 obtained an average of 11,728 reads per sample (± 1 standard deviation: 2962) assigned to
520 family level RDP taxonomy, of which on average 94% (± 1 standard deviation: 4.8%) could
521 be assigned to genus or *Enterobacteriaceae* level taxa.

522 The composition of the bacterial community at family or genus/*Enterobacteriaceae* level
523 was as expected from the known composition of the mock community, and the relative
524 abundances measured in the showed strong, significant correlations with the expected relative
525 abundances of the ZymoBIOMICS™ Microbial Community Standard (family: Pearson
526 $r=0.74$, $p=0.035$; genus/*Enterobacteriaceae*: Pearson $r=0.73$, $p=0.038$), however this was
527 distorted by the over-representation of *Enterobacteriaceae* reads within the dataset, with very
528 poor differentiation of *Escherichia* or *Salmonella* within the taxon. This is consistent with the
529 weaker phylogenetic signal of 16S rRNA genes within these taxa (Fukushima et al. 2002).

530 To provide an insight to the sensitivity of in-field nanopore sequencing to the presence of
531 taxa at low relative abundance, we introduced *Micrococcus luteus* NCTC 2665 genomic
532 DNA to provide 16S rRNA gene copies to provide expected relative abundances in the range
533 0-1.2%. Since the ZymoBIOMICS™ Microbial Community Standard does not contain any
534 *Actinobacteria*, we assumed that reads assigned to its family *Micrococcaceae* or genus
535 *Micrococcus* corresponded to 16S rRNA genes introduced by us. In the three samples of the
536 non-spiked baseline community, only one read was detected assigned to *Micrococcaceae*,
537 however reads assigned to *Micrococcaceae* or *Micrococcus* were detected in two of three
538 samples at 0.4% expected relative abundance, and then in all samples at 0.8-1.2% expected
539 relative abundance. The relative abundance of reads assigned to *Micrococcaceae* and
540 *Micrococcus* were positively significantly correlated with the expected relative abundance

Running title: Extreme Metagenomics

541 (both: Pearson $r=0.66$, $p=0.02$). While future work in developing nanopore-optimized
542 bioinformatics pipelines for 16S rRNA gene analyses could improve reliable assignment of
543 reads to discrete operational taxonomic units, we contend our our analyses highlight the
544 viability of rapid, in-field sequencing of 16S rRNA genes on highly portable MinION devices
545 for comparison of microbial communities.

546 Portable microbiome sequencing: implications for microbial ecology

547 Within this study we describe the application of portable nanopore sequencing for the in-field
548 characterization and comparison of microbial communities. Refinement of our approaches
549 illustrates the potential for generating and analysing sequence data within remote locations
550 without recourse to server-supported bioinformatics, and even in highly resource limited
551 settings.

552 Advances in high throughput sequencing capacity are underpinning continual revolution,
553 revealing revelations about the structure and potential function of microbial communities
554 inhabiting niches throughout every conceivable habitat in the biosphere. Just over a decade
555 ago, Curtis (2006) wrote of the urgent need for microbial ecologists to “go large” and embark
556 upon the high throughput characterization of microbial diversity. Such data-collection
557 initiatives represent essential pre-requisites to the development of mechanistic and predictive
558 insights. Arguably, the vision set out by Curtis (2006) is being accomplished by initiatives
559 such as the Earth Microbiome Project (Thompson et al., 2017), enhancing our coverage of
560 microbial diversity across the planet. However, cataloguing microbial diversity has hitherto
561 been contingent on laboratory-based, high-throughput sequencing platforms, replete with
562 high levels of laboratory infrastructure. This has been at the cost of agility in characterizing
563 and comparing microbial diversity. In-field sequencing using portable MinION sequencers
564 and laptop-based bioinformatics approaches described herein offers the opportunity to regain

Running title: Extreme Metagenomics

565 agility in the characterization of Earth's microbiomes by supporting distributed, at-source
566 DNA sequencing. Loman and Gardy (2017) advocate the merger of human, animal and
567 environmental genomic surveillance through scalable, portable DNA sequencing for digital
568 epidemiology with the goal of achieving a "sequencing singularity". Such a sequencing
569 singularity offers benefits for navigating Earth's microbial diversity in the broader sense also.
570 We contend it is now time to "go small".

571 "Going small" has three conspicuous advantages. Firstly, microbial processes sense and
572 amplify the impacts of environmental change(Vincent, 2010). Understanding the genomic
573 basis of such processes therefore represents a research priority. Parallels may be drawn with
574 the promise of portable sequencing enabled disease diagnosis and surveillance (Gardy and
575 Loman, 2017). For example, within the context of the global cryosphere, rapid warming is
576 creating hitherto unseen opportunities and challenges in the study of microbial diversity.
577 Presently, we lack DNA datasets for over 99.5%(Edwards, 2015; Edwards and Cook, 2015)
578 of Earth's glaciers, which represent Earth's largest freshwater ecosystem (Edwards et al.,
579 2014a) and yet are highly endangered by climatic warming(Hotaling et al., 2017; Milner et
580 al., 2017). Rapid, on-site investigation of microbial interactions with climate warming has the
581 potential to develop baseline data for fragile ecosystems such as glaciers, and to gather
582 observational data as a prelude to hypothesis testing.

583 Secondly, accessing microbial diversity in remote locations, rather than merely its collection
584 and transfer for analysis, also reduces logistical risks. This takes many forms. Sequencing on-
585 site precludes the risk of post-collection changes in community structure, sample degradation
586 or loss in transit *e.g.*(Choo et al., 2015; Hodson et al., 2017). Moreover, investigators gain the
587 flexibility to adjust their sampling campaigns, maximising the value of field campaigns.
588 Furthermore, retrieval of microbial diversity from inaccessible environments requires
589 sophisticated engineering approaches to assure the integrity of recovered samples. Examples

Running title: Extreme Metagenomics

590 include clean subglacial lake access (Siegert et al., 2012; Christner et al., 2014). In-field
591 sequencing offers the possibility of monitoring the recovery of high quality samples, the
592 detection and prevention of contamination, and the optimization of sample capture events
593 within a single field campaign.

594 Thirdly, in-field sequence based characterization and comparison of microbial communities
595 offers the opportunity for distributed characterization of Earth’s microbiomes, thus expanding
596 both our geographical and genomic coverage of microbial diversity. Gilbert et al argue that if
597 Darwin had the technological capacity, he would have used metagenomics in the surveys of
598 biodiversity which underpinned the formulation of the Origin of Species (Darwin, 1859).
599 Indeed if Darwin had been a metagenomic scientist, then *HMS Beagle* would likely have
600 been equipped for in-field sequencing for the discovery of metagenomic diversity.

601 Such considerations are not whimsical. The disparity between the technological constraints
602 on enumerating plant and animal biodiversity and microbial biodiversity underlie a schism
603 between “animal and plant” ecology and “microbial” ecology. This is consequential in that
604 whether microbial equivalents of long-established laws in animal and plant biogeography
605 exist remains a contemporary research question e.g. (Carbonero et al., 2014; van der Gast,
606 2015). We highlight the potential for distributed discovery of Earth’s microbiomes supported
607 by in-field DNA sequencing to underpin a new generation of scientific exploration – and
608 explorers.

609 Conclusions

610 Here we report the use of portable nanopore sequencing to characterize and compare
611 microbial communities while in the field. Our approaches robustly characterize the
612 taxonomic composition of glacial microbial communities using shotgun metagenomics, and
613 permit their comparison by 16S rRNA gene amplicon sequencing. The experiments reported

Running title: Extreme Metagenomics

614 show the versatility of nanopore sequencing approaches for microbiome analyses in a range
615 of field settings, and the coherence of data produced with established approaches for
616 investigating microbial diversity. Continued development in the field of nanopore
617 sequencing, for example the use of lyophilized reagents and ambient-temperature stable
618 nanopore flow cells for sequencing with on-laptop analytic approaches raises the prospect of
619 highly agile characterization of Earth's microbiomes at source.

620 **Acknowledgements**

621 AE is supported by a Leverhulme Research Fellowship (RF-2017-652), a Royal Geographic
622 Society Walters Kundert Arctic Fellowship and acknowledges the support of the HEFCW
623 Capital Research Infrastructure Fund and BBSRC funded facilities required for the analyses
624 described herein. NERC (NE/K000942/1) supported alpine fieldwork. JMC acknowledges
625 the Rolex Enterprise Award which supported Greenland fieldwork. ARD is supported by a
626 Life Sciences Research Network Wales PhD scholarship while SR is supported by a Sêr
627 Cymru Fellowship. AE, JMC and TD gratefully thank Eddie Frost and Jeffery Garriock for
628 their support and *perseverance* in the face of technical challenges on Greenland.

629 **Author contribution**

630 AE: Conceived study, conducted fieldwork, sequencing, analysis, wrote manuscript. ARD:
631 Developed DNA extraction method, conducted fieldwork. SMN, LAJM, AS and SR:
632 bioinformatics analysis and support; BS, JMC, TD and AJH: conducted fieldwork. All
633 authors contributed to and commented upon the manuscript.

634 **Conflicting interest statement**

Running title: Extreme Metagenomics

635 AE has received financial support to attend and present elements of this work at the
636 Nanopore Community Meetings 2016 and 2017 and free reagents for outreach work from
637 Oxford Nanopore Technologies, Ltd.

638

Table 1: Summary of sequencing experiments

Experiment	Analysis type	Study Site	DNA extraction	Library kit
Svalbard RAD001	Shotgun	Foxfonna	Mod-Powersoil	RAD001
Greenland RAD001	Shotgun	Russell glacier	Mod-Powersoil	RAD001
Austria RAD003	Shotgun	Rotmoosferner	Mod-Powersoil	RAD003
Austria LRK001	Shotgun	Rotmoosferner	Mod-Powersoil	LRK001
Mock LRK001	Shotgun	Lab	N/A	LRK001
Svalbard 16S	16S rRNA gene	Vestre Brøggerbreen	Mod-Powersoil	RAB201
Arctic-Alps 16S	16S rRNA gene	Lab	Mod-Powersoil	RAB201
Mock-Spike 16S	16S rRNA gene	Lab	N/A	RAB201

639

640

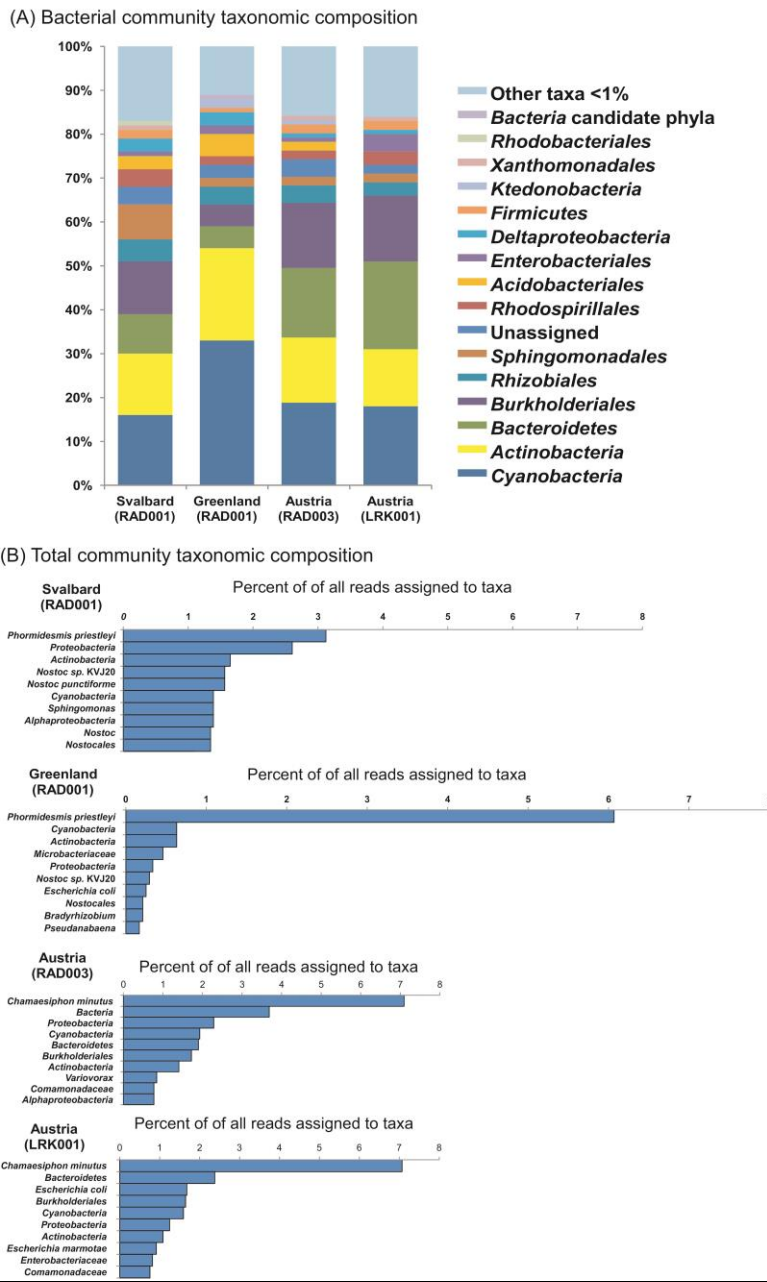
641

642

643

Running title: Extreme Metagenomics

644 **Figures**

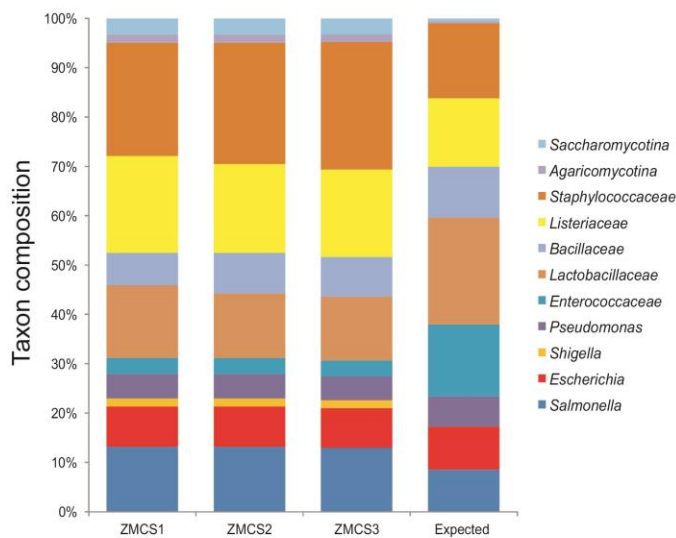


645

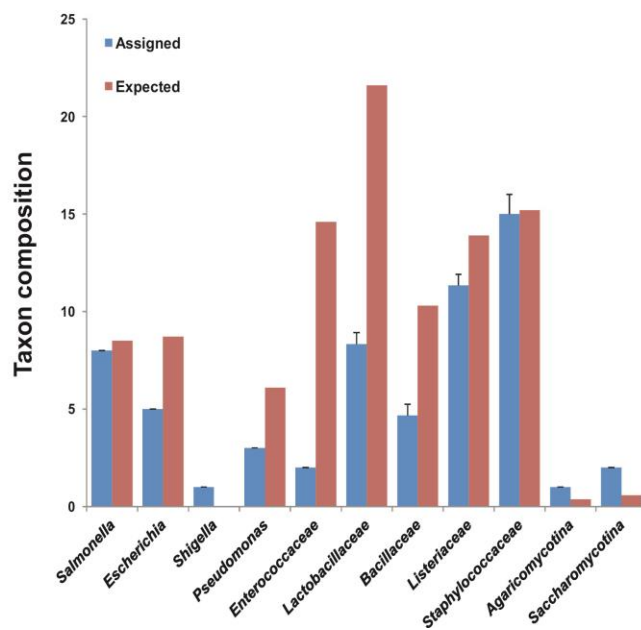
646 **Figure 1:** In field metagenome sequencing reveals taxonomic distribution of cryoconite
 647 microbiota. (A): shows bacterial taxa present at abundance greater than 1% of the bacterial
 648 community while (B) shows the top ten taxa as a percentage of reads assigned to any taxon.
 649 Svalbard and Greenland datasets were generated in the field, while Austria datasets were
 650 generated using field protocols to compare aqueous (RAD003) and lyophilised library
 651 preparation (LRK001) approaches.

Running title: Extreme Metagenomics

(A) Taxonomic composition of ZymoBIOMICS(TM) Microbial Community Standard obtained by lyophilized field kit nanopore sequencing



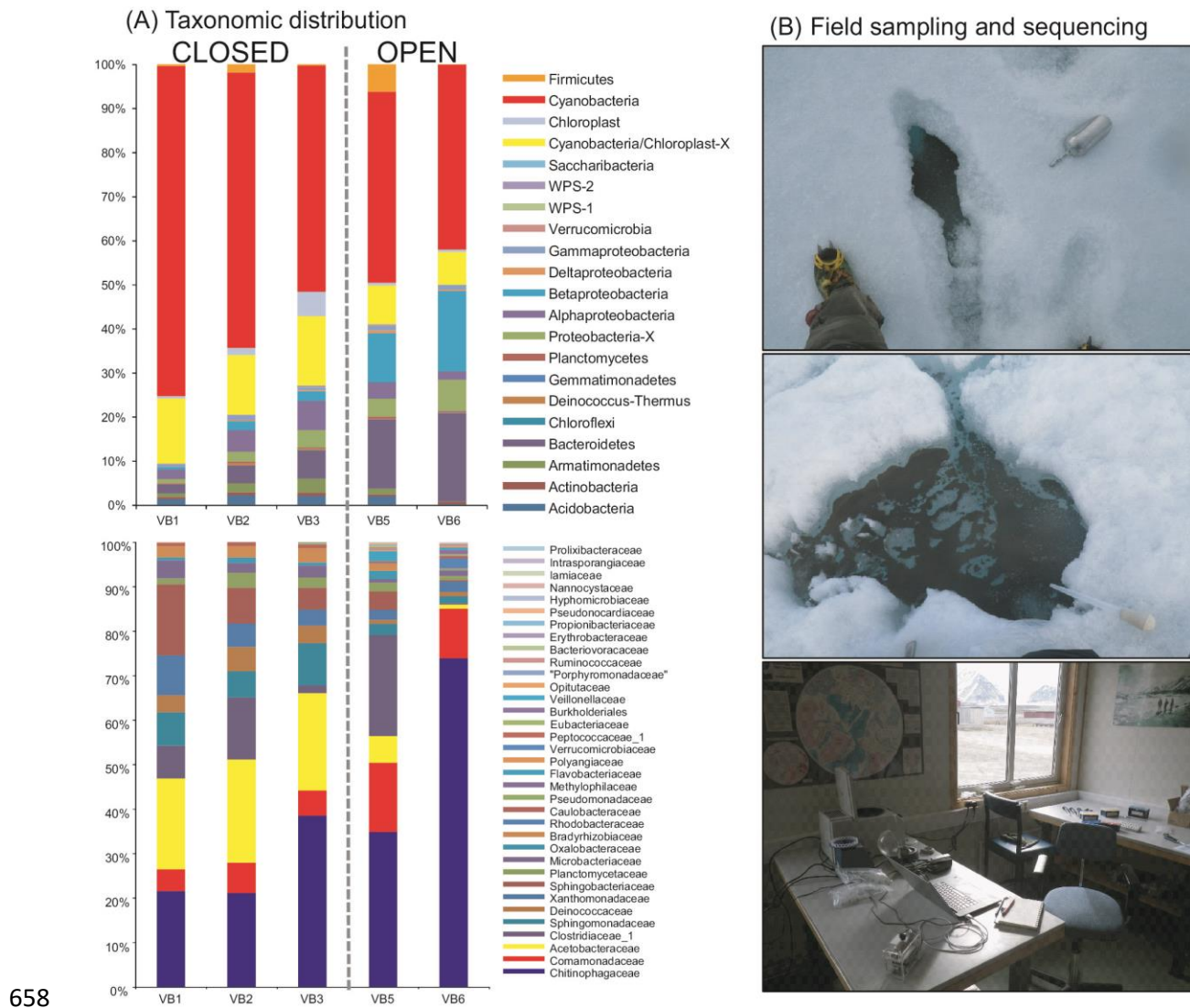
(B) Comparison of taxa present in the ZymoBIOMICS (TM) Microbial Community Standard detected by lyophilized field kit nanopore sequencing to expected relative abundance



652

653 **Figure 2:** Benchmarking of in-field nanopore sequencing using lyophilized reagents. The
654 ZymoBIOMICS Microbial Community Standard was sequenced in triplicate (ZMCS1-3) to
655 compare against its expected composition and relative abundance profile. (A) shows the
656 overall profile of all taxa detected at >1% relative abundance while (B) compares the relative
657 abundance of assigned taxa against the expected composition.

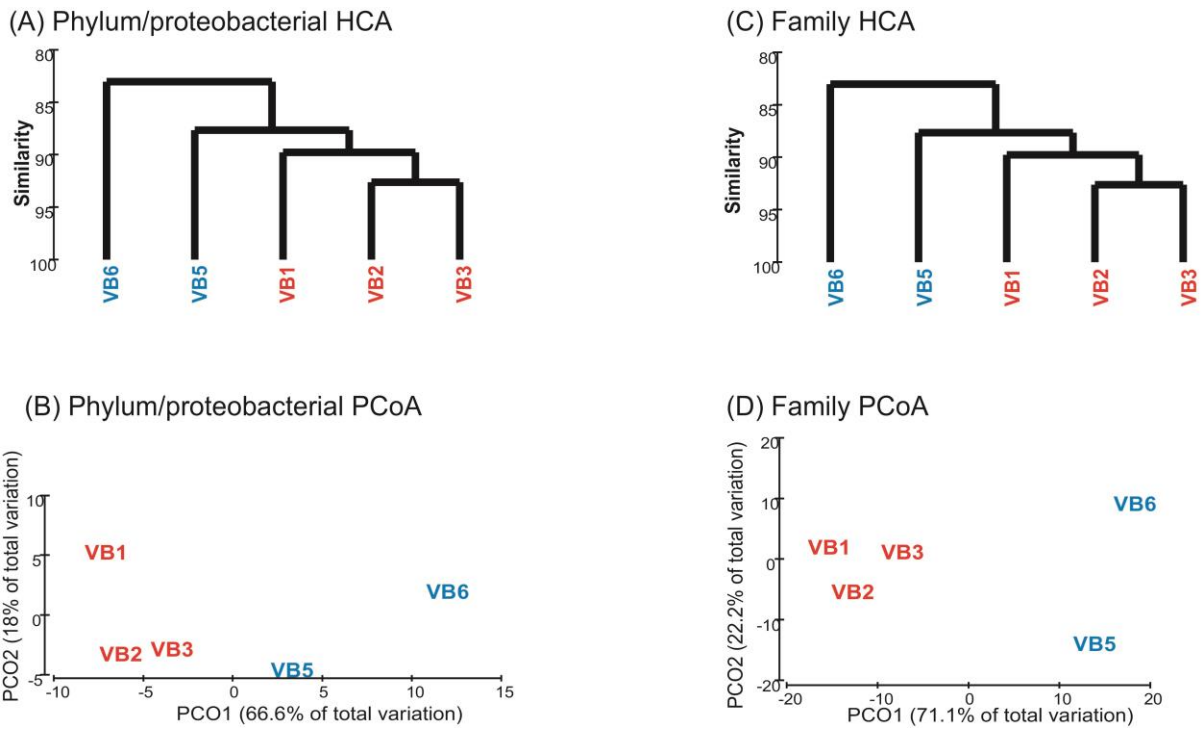
Running title: Extreme Metagenomics



658

659 **Figure 3:** In-field 16S rRNA gene amplicon sequencing of cryoconite bacterial communities
 660 showing (A) phylum and family level taxonomic distribution of (B) open and closed
 661 cryoconite holes sampled on Vestre Brøggerbreen. The top image shows a “closed”
 662 cryoconite hole where the snow and superimposed ice cover has been displaced to reveal the
 663 hole, while the middle image shows a seasonally-open cryoconite hole. The bottom image
 664 shows the arrangement of equipment for DNA extraction, 16S rRNA gene PCR and nanopore
 665 sequencing in a field lab.

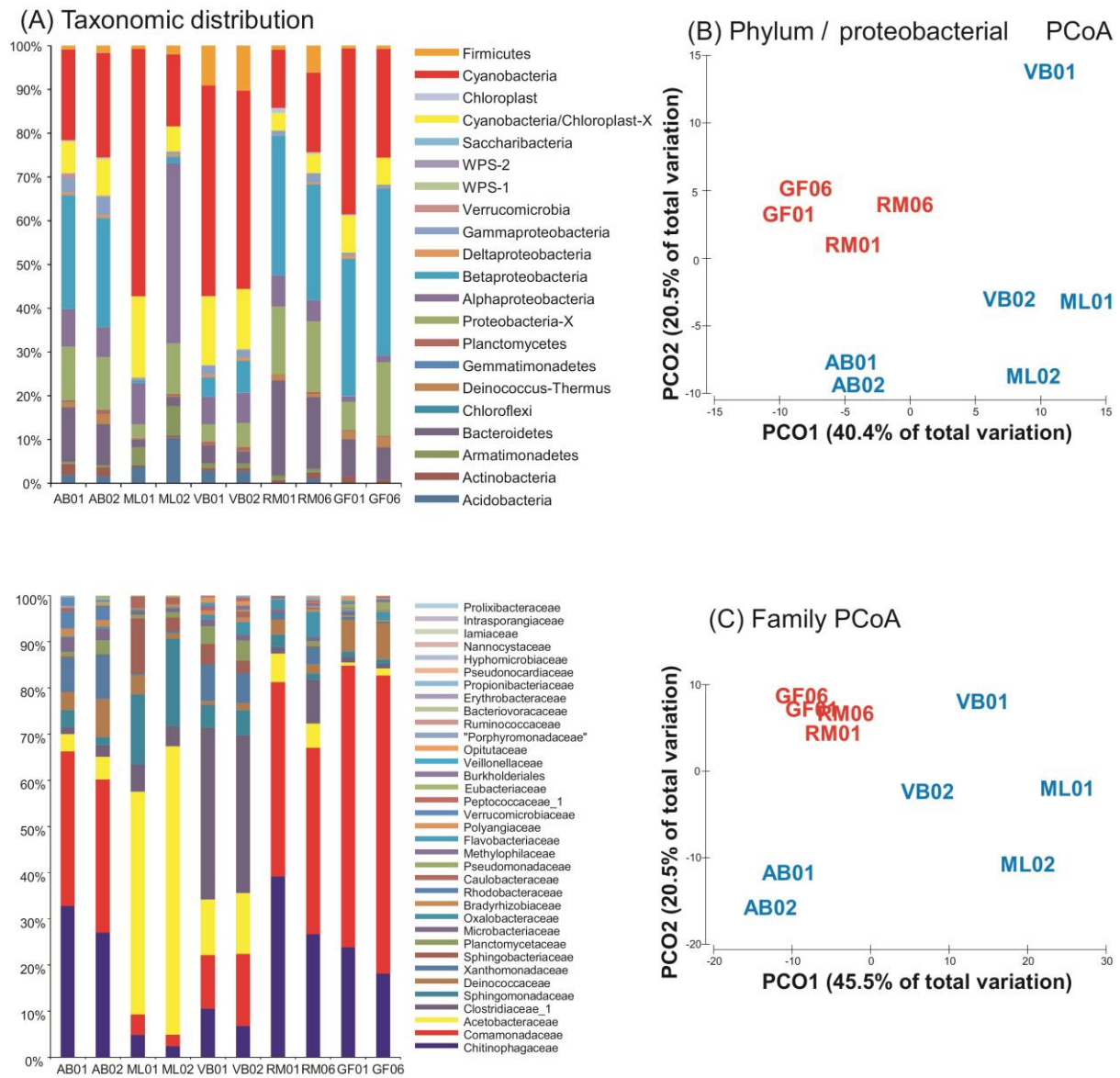
Running title: Extreme Metagenomics



666

667 **Figure 4:** multivariate discrimination of cryoconite bacterial communities revealed by in-
668 field 16S rRNA gene amplicon sequencing. Analyses are performed with data aggregated to
669 phylum/proteobacterial class (A-B) or family-level taxa (C-D) with Hierarchical Cluster
670 Analysis (HCA, subpanels A, C,) or Principal Coordinates Analysis (PCoA, B,D,). Closed
671 holes (VB1-3) and open holes (VB5-6) are ordinated by multivariate analysis of of fourth-
672 root transformed Bray-Curtis distances of phylotype relative abundances.

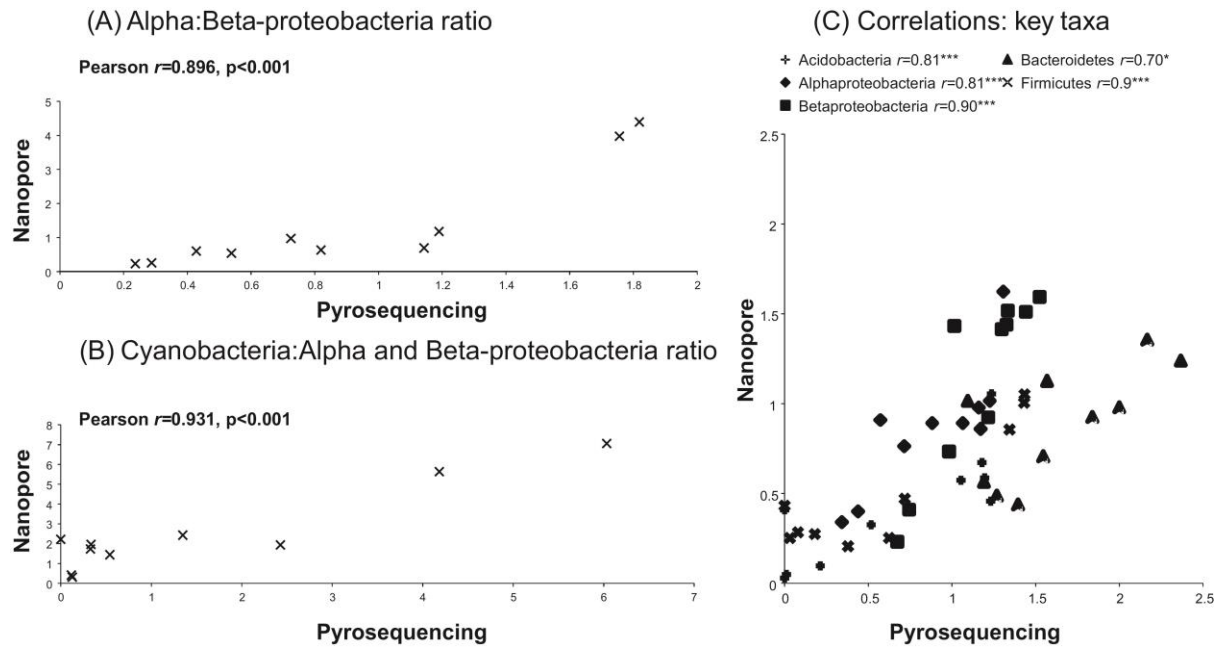
Running title: Extreme Metagenomics



673

674 **Figure 5:** Benchmarking nanopore 16S rRNA gene amplicon sequencing of cryoconite
 675 bacterial communities by comparison with laboratory-generated data. (A) Phylum and family
 676 level taxonomic distribution of cryoconite bacterial communities and (B) phylum and (C)
 677 family level taxon distributions used for principal coordinates analysis of fourth-root
 678 transformed Bray-Curtis distances of phylotype relative abundances discriminate between
 679 Arctic and alpine cryoconite communities. Arctic glaciers (blue, AB, ML, VB) and Alpine
 680 glaciers (red, GF, RM) are clearly ordinated.

Running title: Extreme Metagenomics



681

682 **Figure 6:** Benchmarking nanopore 16S rRNA gene amplicon sequencing of cryoconite

683 bacterial communities by comparison with laboratory-generated data. Correlation of log

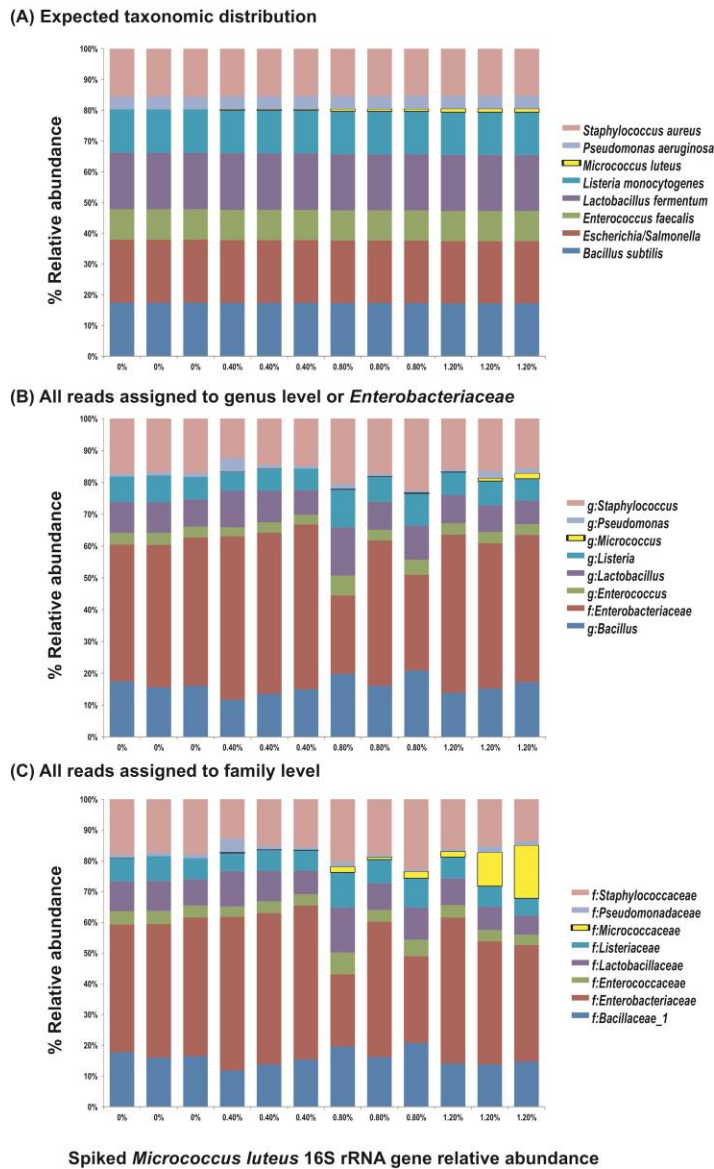
684 relative abundances between (A) the ratio of *Alphaproteobacteria* and *Betaproteobacteria*

685 (B) *Cyanobacteria* to *Alphaproteobacteria*:*Betaproteobacteria* and (C) key taxonomic

686 groups revealed using nanopore and pyro- sequencing of 16S rRNA genes. Positive,

687 significant or highly significant Pearson r correlations are observed for each taxon.

Running title: Extreme Metagenomics



689 **Figure 7:** Benchmarking nanopore 16S rRNA gene amplicon sequencing by sequencing
690 mock communities using in-field protocols. Panel (A) shows the expected taxonomic
691 distribution of The ZymoBIOMICS Microbial Community Standard for each sample, spiked
692 in triplicate with *Micrococcus luteus* NCTC2665 genomic DNA to afford theoretical relative
693 abundances of 16S rRNA genes in the range of 0-1.2% of the total bacterial community.
694 Panel (B) shows shows the observed data where all reads assigned to genus level (excepting
695 the poorly resolvable *Enterobacteriaceae* while Panel (C) shows all reads at the family level.
696 The expected and observed *Micrococcus luteus* spike is highlighted in yellow.

697 **REFERENCES**

- 698 Anesio, A.M., Hodson, A.J., Fritz, A., Psenner, R., and Sattler, B. (2009). High microbial
699 activity on glaciers: importance to the global carbon cycle. *Global Change Biology*
700 15(4), 955-960. doi: 10.1111/j.1365-2486.2008.01758.x.
- 701 Anesio, A.M., Sattler, B., Foreman, C., Telling, J., Hodson, A., Tranter, M., et al. (2010).
702 Carbon fluxes through bacterial communities on glacier surfaces. *Annals of*
703 *Glaciology* 51(56), 32-40.
- 704 Brown, B.L., Watson, M., Minot, S.S., Rivera, M.C., and Franklin, R.B. (2017). MinION™
705 nanopore sequencing of environmental metagenomes: a synthetic approach.
706 *GigaScience* 6(3), 1-10.
- 707 Calus, S.T., Ijaz, U.Z., and Pinto, A.J. (2018). NanoAmpli-Seq: A workflow for amplicon
708 sequencing from mixed microbial communities on the nanopore sequencing platform.
709 *bioRxiv*. doi: 10.1101/244517.
- 710 Cameron, K.A., Hodson, A.J., and Osborn, A.M. (2012). Structure and diversity of bacterial,
711 eukaryotic and archaeal communities in glacial cryoconite holes from the Arctic and
712 the Antarctic. *FEMS Microbiology Ecology* 82, 254-267. doi: 10.1111/j.1574-
713 6941.2011.01277.x.
- 714 Carbonero, F., Oakley, B.B., and Purdy, K.J. (2014). Metabolic flexibility as a major
715 predictor of spatial distribution in microbial communities. *PloS One* 9(1), e85105.
- 716 Castro-Wallace, S.L., Chiu, C.Y., John, K.K., Stahl, S.E., Rubins, K.H., McIntyre, A.B.R., et
717 al. (2017). Nanopore DNA Sequencing and Genome Assembly on the International
718 Space Station. *Scientific Reports* 7(1), 18022. doi: 10.1038/s41598-017-18364-0.
- 719 Choo, J.M., Leong, L.E.X., and Rogers, G.B. (2015). Sample storage conditions significantly
720 influence faecal microbiome profiles. *Scientific Reports* 5, 16350. doi:
721 10.1038/srep16350.
- 722 Christner, B.C., Priscu, J.C., Achberger, A.M., Barbante, C., Carter, S.P., Christianson, K., et
723 al. (2014). A microbial ecosystem beneath the West Antarctic ice sheet. *Nature*
724 512(7514), 310-313. doi: 10.1038/nature13667
- 725 Cook, J.M., Edwards, A., Bulling, M., Mur, L.A., Cook, S., Gokul, J.K., et al. (2016).
726 Metabolome-mediated biocryomorphic evolution promotes carbon fixation in
727 Greenlandic cryoconite holes. *Environmental microbiology* 18(12), 4674-4686.

Running title: Extreme Metagenomics

- 728 Cook, J.M., Edwards, A., Takeuchi, N., and Irvine-Fynn, T.D.L. (2015). Cryoconite: the dark
729 biological secret of the cryosphere. *Progress in Physical Geography* in press.
- 730 Curtis, T. (2006). Microbial ecologists: it's time to 'go large'. *Nature Reviews Microbiology*
731 4(7), 488-488.
- 732 Darwin, C. (1859). *On the Origin of Species*. London: John Murray.
- 733 Edgar, R. (2016). SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
734 sequences. *bioRxiv*. doi: 10.1101/074161.
- 735 Edwards, A. (2015). Coming in from the cold: potential microbial threats from the terrestrial
736 cryosphere. *Frontiers in Earth Science* 3. doi: 10.3389/feart.2015.00012.
- 737 Edwards, A., Anesio, A.M., Rassner, S.M., Sattler, B., Hubbard, B., Perkins, W.T., et al.
738 (2011). Possible interactions between bacterial diversity, microbial activity and
739 supraglacial hydrology of cryoconite holes in Svalbard. *The ISME Journal* 5(1), 150-
740 160. doi: DOI 10.1038/ismej.2010.100.
- 741 Edwards, A., and Cook, S. (2015). Microbial dynamics in glacier forefield soils show
742 succession is not just skin deep. *Molecular Ecology* 24(5), 963-966. doi:
743 10.1111/mec.13098.
- 744 Edwards, A., Debbonaire, A.R., Sattler, B., Mur, L.A., and Hodson, A.J. (2016). Extreme
745 metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N.
746 *bioRxiv*, 073965.
- 747 Edwards, A., Irvine-Fynn, T., Mitchell, A.C., and Rassner, S.M.E. (2014a). A germ theory
748 for glacial systems? *Wiley Interdisciplinary Reviews: Water* 1(4), 331-340. doi:
749 10.1002/wat2.1029.
- 750 Edwards, A., Mur, L.A.J., Girdwood, S.E., Anesio, A.M., Stibal, M., Rassner, S.M.E., et al.
751 (2014b). Coupled cryoconite ecosystem structure–function relationships are revealed
752 by comparing bacterial communities in alpine and Arctic glaciers. *FEMS*
753 *Microbiology Ecology* 89(2), 222-237. doi: 10.1111/1574-6941.12283.
- 754 Edwards, A., Pachebat, J.A., Swain, M., Hegarty, M., Hodson, A., Irvine-Fynn, T.D.L., et al.
755 (2013a). A metagenomic snapshot of taxonomic and functional diversity in an alpine
756 glacier cryoconite ecosystem. *Environmental Research Letters* 8(3), 035003.
- 757 Edwards, A., Pachebat, J.A., Swain, M., Hegarty, M., Hodson, A.J., Irvine-Fynn, T., et al.
758 (2013b). A metagenomic snapshot of taxonomic and functional diversity in an alpine
759 glacier cryoconite ecosystem. *Environmental Research Letters* 8(3), 035003.

Running title: Extreme Metagenomics

- 760 Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The Microbial Engines That Drive
761 Earth's Biogeochemical Cycles. *Science* 320(5879), 1034-1039. doi:
762 10.1126/science.1153213.
- 763 Fukushima, F., Kakinuma, K., and Kawaguchi, R. (2002). Phylogenetic analysis of
764 *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene
765 sequence. *Journal of Clinical Microbiology* 40(8):2779-2785.
- 766 Gardy, J.L., and Loman, N.J. (2017). Towards a genomics-informed, real-time, global
767 pathogen surveillance system. *Nature Reviews Genetics* 19, 9. doi:
768 10.1038/nrg.2017.88.
- 769 Gilbert, J.A., O'Dor, R., King, N., and Vogel, T.M. (2011). The importance of metagenomic
770 surveys to microbial ecology: or why Darwin would have been a metagenomic
771 scientist. *Microbial Informatics and Experimentation* 1(1), 5.
- 772 Gokul, J.K., Hodson, A.J., Saetnan, E.R., Irvine-Fynn, T.D., Westall, P.J., Detheridge, A.P.,
773 et al. (2016). Taxon interactions control the distributions of cryoconite bacteria
774 colonizing a High Arctic ice cap. *Molecular Ecology* 25(15), 3752-3767.
- 775 Goordial, J., Altshuler, I., Hindson, K., Chan-Yam, K., Marcoléfas, E., and Whyte, L.G.
776 (2017). In Situ Field Sequencing and Life Detection in Remote (79°26'N) Canadian
777 High Arctic Permafrost Ice Wedge Microbial Communities. *Frontiers in*
778 *Microbiology* 8(2594). doi: 10.3389/fmicb.2017.02594.
- 779 Hodson, A., Nowak, A., Cook, J., Sabacka, M., Wharfe, E., Pearce, D., et al. (2017).
780 Microbes influence the biogeochemical and optical properties of maritime Antarctic
781 snow. *Journal of Geophysical Research: Biogeosciences*.
- 782 Hotaling, S., Hood, E., and Hamilton, T.L. (2017). Microbial ecology of mountain glacier
783 ecosystems: Biodiversity, ecological connections, and implications of a warming
784 climate. *Environmental Microbiology*.
- 785 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., et al.
786 (2016). A new view of the tree of life. *Nature Microbiology* 1, 16048.
- 787 Johnson, S.S., Zaikova, E., Goerlitz, D.S., Bai, Y., and Tighe, S.W. (2017). Real-Time DNA
788 Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer.
789 *Journal of Biomolecular Techniques : JBT* 28(1), 2-7. doi: 10.7171/jbt.17-2801-009.
- 790 Kerkhof, L.J., Dillon, K.P., Häggblom, M.M., and McGuinness, L.R. (2017). Profiling
791 bacterial communities by MinION sequencing of ribosomal operons. *Microbiome*
792 5(1), 116.

Running title: Extreme Metagenomics

- 793 Klein, D.A. (2015). Partial Formalization: An Approach for Critical Analysis of Definitions
794 and Methods Used in Bulk Extraction-Based Molecular Microbial Ecology. *Open*
795 *Journal of Ecology* 5(08), 400.
- 796 Langford, H., Hodson, A., Banwart, S., and Bøggild, C. (2010). The microstructure and
797 biogeochemistry of Arctic cryoconite granules. *Annals of Glaciology* 51(56), 87-94.
- 798 Li, C., Chng, K.R., Boey, E.J.H., Ng, A.H.Q., Wilm, A., and Nagarajan, N. (2016). INC-Seq:
799 accurate single molecule reads using nanopore sequencing. *GigaScience* 5(1), 34.
- 800 Lutz, S., Anesio, A.M., Edwards, A., and Benning, L.G. (2016). Linking microbial diversity
801 and functionality of arctic glacial surface habitats. *Environmental Microbiology*.
- 802 Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for
803 metagenomics with Kaiju. *Nature communications* 7.
- 804 Milner, A.M., Khamis, K., Battin, T.J., Brittain, J.E., Barrand, N.E., Füreder, L., et al. (2017).
805 Glacier shrinkage driving global changes in downstream systems. *Proceedings of the*
806 *National Academy of Sciences* 114(37), 9770-9778.
- 807 Musilova, M., Tranter, M., Bennett, S.A., Wadham, J.L., and Anesio, A. (2015). Stable
808 microbial community composition on the Greenland Ice Sheet. *Frontiers in*
809 *Microbiology* 6. doi: 10.3389/fmicb.2015.00193.
- 810 Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., et al. (2016).
811 Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530(7589),
812 228-232.
- 813 Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., et al.
814 (2013). Insights into the phylogeny and coding potential of microbial dark matter.
815 *Nature* 499(7459), 431-437. doi: 10.1038/nature12352
- 816 Rutter, N., Hodson, A., Irvine-Fynn, T., and Solas, M.K. (2011). Hydrology and
817 hydrochemistry of a deglaciating high-Arctic catchment, Svalbard. *Journal of*
818 *Hydrology* 410(1-2), 39-50. doi: DOI 10.1016/j.jhydrol.2011.09.001.
- 819 Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., et al. (2014).
820 Reagent and laboratory contamination can critically impact sequence-based
821 microbiome analyses. *BMC Biology* 12(1), 87.
- 822 Segawa T., Yonezawa T., Edwards A., Akiyoshi, A., Tanaka, S., Uetake, J., Irvine-Fynn,
823 T.D.L., Fukui, K., Li, Z and Takeuchi, N. et al. (2017). Biogeography of cryoconite
824 forming cyanobacteria on polar and Asian glaciers. *Journal of Biogeography*
825 44:2849–2861.

Running title: Extreme Metagenomics

- 826 Schmidt, T.M., DeLong, E., and Pace, N. (1991). Analysis of a marine picoplankton
827 community by 16S rRNA gene cloning and sequencing. *Journal of bacteriology*
828 173(14), 4371-4378.
- 829 Siegert, M.J., Clarke, R.J., Mowlem, M., Ross, N., Hill, C.S., Tait, A., et al. (2012). Clean
830 access, measurement, and sampling of Ellsworth Subglacial Lake: a method for
831 exploring deep Antarctic subglacial lake environments. *Reviews of Geophysics* 50(1).
- 832 Stibal, M., Schostag, M., Cameron, K.A., Hansen, L.H., Chandler, D.M., Wadham, J.L., et al.
833 (2015). Different bulk and active bacterial communities in cryoconite from the margin
834 and interior of the Greenland ice sheet. *Environmental microbiology reports* 7(2),
835 293-300.
- 836 Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., et al. (2017).
837 A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*
838 551(7681).
- 839 Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al.
840 (2004). Community structure and metabolism through reconstruction of microbial
841 genomes from the environment. *Nature* 428(6978), 37-43.
- 842 Uetake, J., Tanaka, S., Segawa, T., Takeuchi, N., Nagatsuka, N., Motoyama, H., et al. (2016).
843 Microbial community variation in cryoconite granules on Qaanaaq Glacier, NW
844 Greenland. *FEMS microbiology ecology* 92(9), fiw127.
- 845 van der Gast, C.J. (2015). Microbial biogeography: the end of the ubiquitous dispersal
846 hypothesis? *Environmental Microbiology* 17(3), 544-546. doi: 10.1111/1462-
847 2920.12635.
- 848 Vincent, W.F. (2010). Microbial ecosystem responses to rapid climate change in the Arctic.
849 *The ISME Journal* 4(9), 1087-1090.
- 850 Watson, M., Thomson, M., Risse, J., Talbot, R., Santoyo-Lopez, J., Gharbi, K., et al. (2015).
851 poRe: an R package for the visualization and analysis of nanopore sequencing data.
852 *Bioinformatics* 31(1), 114-115.
- 853 Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T., et al. (2016).
854 Challenges in microbial ecology: building predictive understanding of community
855 function and dynamics. *The ISME Journal*. 10 11:2557-2568.
- 856 Young, M., Artsatbanov, V., Beller, H.R., Chandra, G., Chater, K.F., Dover, L.G., et al.
857 (2010). Genome sequence of the Fleming strain of *Micrococcus luteus*, a simple free-
858 living actinobacterium. *Journal of Bacteriology*. 192(3), 841-60.