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

Arwyn Edwards[1,2]*, Aliyah R. Debbonaire[1,2], Samuel M. Nicholls[1,2,3] Sara M.E. Rassner[2,4], Birgit Sattler[5], Joseph M. Cook[6], Tom Davy[1,2], Luis A.J. Mur[1], Andrew J. Hodson[7]

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

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

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

ABSTRACT

In the field of observation, chance favours only the prepared mind (Pasteur). Impressive developments in genomics have led microbiology to its third "Golden Age". However, conventional metagenomics strategies necessitate retrograde transfer of samples from extreme or remote environments for later analysis, rendering the powerful insights gained retrospective in nature, striking a contrast with Pasteur's dictum. Here we implement highly portable USB-based nanopore DNA sequencing platforms coupled with field-adapted environmental DNA extraction, rapid sequence library generation and off-line analyses of shotgun metagenome and 16S ribosomal RNA gene amplicon profiles to characterize microbiota dwelling within cryoconite holes upon Svalbard glaciers, the Greenland Ice Sheet and the Austrian Alps. We show in-field nanopore sequencing of metagenomes captures taxonomic composition of supraglacial microbiota, while 16S rRNA gene amplicon sequencing resolves bacterial community responses to habitat changes. We benchmark the capability of in-field microbiome sequencing to characterize microbial communities by comparison of nanopore data with prior Illumina metagenomic data and 16S rRNA gene V1-V3 pyrosequencing from the same samples, demonstrating a high level of coherence between profiles obtained from nanopore sequencing and laboratory based sequencing approaches. Ultimately, in-field sequencing potentiated by nanopore devices raises the prospect of enhanced agility in exploring Earth's most remote microbiomes.

INTRODUCTION

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

Sampling

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

In-field DNA extraction

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

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

beating was conducted using a Vortex Genie 2 (Scientific Industries, Inc) fitted with a MoBio (MoBio, Inc.) tube adapter or a IKA Works, Inc MS2 S8 Minishaker. Since the PowerSoil® protocol typically uses benchtop centrifuges capable of 15-17,000 ×g, the use of a Gilson, Ltd GmCLab Microcentrifuge capable of generating <2900 ×g required elongated spin times of four minutes for protein precipitation following the addition of solution C2 and inhibitor removal following the addition of solution C3. For all other centrifugation steps, spin times of one minute were sufficient. To increase DNA concentration, DNA was eluted in 70 µL solution C6, after a five minute incubation at room temperature. Two microlitre aliquots of each extract were than immediately quantified using Qubit 2.0 Fluorimeter dsDNA High Sensitivity assays (Invitrogen, Ltd). In the case of Greenland metagenomics, fluorimetry was performed with the Qubit device powered using a portable power-pack (PowerAdd, Inc.)

Svalbard: shotgun metagenomics experiment, Foxfonna ice cap.

On 18 August 2016, cryoconite debris from four cryoconite holes was collected as described above on the un-named outlet glacier of Foxfonna ice cap (78˚ 08'N, 16˚ 07' E; (Rutter et al., 2011; Gokul et al., 2016).) Cryoconite was thawed the following day and DNA extracted as described. Equimolar pools of cryoconite DNA provided 185 ng DNA used for a transposase-based 1D rapid sequencing library preparation exactly as specified by Oxford Nanopore Technologies, Ltd (Rapid Sequencing of genomic DNA for the minION device using SQK-RAD001) using a PCR cycler for the 30°C and 75°C incubation steps in a Hybaid Omni-E PCR cycler.

1D sequencing was directly performed using a R9 flow cell (Oxford Nanopore Technologies, Ltd) loaded conventionally in 150 µL sequencing mix to a MinION Mk 1 device run using MinKNOW 1.0.5 (Oxford Nanopore Technologies, Ltd). Sequencing and read processing were perfomed on a Lenovo® ThinkPad® X1 Carbon running Windows 10 with Intel®

Core™ i7-5600U Processor, 8GB RAM and a 512GB solid state drive. 1D basecalling was performed remotely via the Oxford Nanopore Technologies, Ltd. Metrichor platform. Template strand nanopore reads in .fast5 format were converted to fasta using the pore_parallel GUI of poRe 0.17 (Watson et al., 2015) running in R 3.3.1.

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

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

The library (75 µL) was loaded in the SpotON port of a R9.4 flow cell (Oxford Nanopore Technologies, Ltd) run using a MinION MK1 device. Due to the cold (<0°C) ambient temperatures, to maintain optimal sequencing temperature, the MinION device was enclosed in an insulated cool bag during sequencing (Tesco, Ltd). Sequencing and read processing were perfomed on a Dell XPS15 laptop running Windows 10 Intel® Core™ i7-6700HQ Processor, 32GB RAM and a 1GB solid state drive, running on internal battery. Sequencing was controlled using MinKNOW 1.6.11 (modified for offline use, courtesy Oxford Nanopore Technologies, Ltd) and 1D basecalling performed locally with Albacore v1.1.0 (Oxford Nanopore Technologies, Ltd).

Benchmarking: shotgun metagenomics experiment, Rotmoosferner, Austrian Alps.

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

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

Storage and transportation of molecular biology reagents and sequencing flow cells at cold temperatures represents a critical limitation to the utility of nanopore sequencing for characterizing microbial communities in remote locations. We therefore tested lyophilized versions of the rapid library kit and ambient-shipped R9.4 flow cells provided by Oxford Nanopore Technologies Ltd. in a repetition of the above experiment. The experiment was performed without laboratory resources. Briefly, 400 ng of community DNA in a 10 µL volume was used to suspend lyophilized fragmentation mix, with incubation at room temperature with inactivation and immersion in hot water stored in an insulated mug for 80°C inactivation for one minute each. The resulting solution was incubated with rapid adapter for

five minutes before gentle mixing with 65 µL priming buffer and flow cell loading before sequencing as described above.

Bioinformatics for shotgun metagenomics

For all experiments, sequencing reads were analysed using Kaiju(Menzel et al., 2016), a taxonomic classifier which is offered as a webserver (http://kaiju.binf.ku.dk.) or as a standalone version. Kaiju seeks protein-level matches in all possible reading frames using the Burrows-Wheeler transform and has been demonstrated to be rapid yet sensitive (Menzel et al., 2016). We report Kaiju analyses conducted using the NCBI *nr+euk* database (version: 05 May 2017) containing 103 million protein sequences from microbes, viruses and selected microbial eukaryotes with low complexity filtering and in *greedy* mode (Minimum match: 11; minimum match score: 75; allowed mismatches: 5). Outputs from Kaiju were visualized as Krona plots. Taxon path and Kaiju output files are available as additional files 1-10. Nanopore sequencing data from in-field metagenomes are available at the European Nucleotide Archive (ENA) accessions ERR2264275 - ERR2264278.

Svalbard: 16S ribosomal RNA gene sequencing, Vestre Brøggerbreen

On 10 July 2017, cryoconite debris was collected from six cryoconite holes on Vestre Brøggerbreen (78° 54’7 N, 011°43’8 E) on Svalbard. Three cryoconite holes (VB1-3) were covered by snow and superimposed ice while three cryoconite holes (VB4-6) were exposed. To test the hypothesis that the presence or absence of snow/ice cover incurs changes in the bacterial community structure, 16S ribosomal RNA gene sequencing was conducted. Samples were transferred to the NERC Arctic Research Station Ny Ålesund within three hours and DNA extracted and quantified as described above within the field lab.

Bacterial 16S rRNA genes were amplified from 50 ng DNA per sample diluted to 10 µL nuclease free water in 50 µL with 1 × LongAmp Taq master mix (New England Biolabs, Inc), 1 µL 16S rRNA gene barcoded primer (ONT-RAB 201, Oxford Nanopore Technologies, Ltd) and 14 µL nuclease free water. Each sample was allocated to a barcoded primer. A no template control (nuclease free water) and extract control (blank extract) were amplified and sequenced in parallel. PCR was conducted as specified in the Oxford Nanopore Technologies protocol for rapid, barcoded sequencing of 16S rRNA genes (ONT-RAB201, Oxford Nanopore Technologies, Ltd) with the modification that 30 cycles of PCR were performed. PCR was performed using a 8 well MiniPCR device (Amplyus, Inc.) controlled by a Windows laptop. The resulting PCR products were purified using magnetic beads (Ampure XP, Beckman Coulter, Ltd) and eluted in 10 µL buffer (10 mM Tris, 50 mM NaCl, pH 8.0) prior to Qubit quantification. Equimolar quantities of each barcode were pooled and sequenced using a R9.4 flow cell as described above for the *Greenland* experiment prior to 1D basecalling and barcode demultiplexing in Albacore v1.10.

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

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

Ten samples (Svalbard: three pairs of samples from Austre Brøggerbreen [AB], Midtre Lovénbreen [ML] and Vestre Brøggerbreen [VB]. Austria; two pairs of samples from

Gaisbergferner[GB] and Rotmoosferner[RM]) archived at -80°C since collection were used to generate fresh DNA extracts as described above. Bacterial 16S rRNA genes were amplified and sequenced exactly as described for the Svalbard 16S ribosomal RNA gene sequencing experiment, with the exception that a 96 well PCR cycler (G-Storm Direct, Ltd) was used for amplification and Albacore v2.02 used for basecalling and demultiplexing by barcode. Samples were assigned an individual barcode primer, with co-amplification and sequencing of no-template and extract controls.

16S rRNA gene data analysis

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

The number of reads assigned per taxon were counted in MS Excel and the relative abundance of reads per phylum (or proteobacterial class) and family used for separate downstream analysis. Multivariate analyses of bacterial community structure using Bray-Curtis distances of fourth root transformed taxon relative abundance data were performed in Primer-6.1.12 & Permanova+1.0.2 (Primer-E, Ltd, Ivybridge, UK). Principal Coordinates Analysis (PCoA) and group-average Hierarchical Cluster Analysis (HCA) were used as

unsupervised data exploration, while hypotheses were tested using PERMANOVA with unrestricted permutation of raw data and 9,999 ordinations. Reads from the implementation and benchmarking experiments are available from ENA at ERR2264279 - ERR2264283 and ERR2264284 - ERR2264293 respectively, and the taxonomic assignments for each read are presented as additional file 11.

RESULTS AND DISCUSSION

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

In field metagenomics of Arctic cryoconite

Here, we report the generation of two metagenomes from Arctic cryoconite by in-field nanopore sequencing, and the taxonomic classification of microbial diversity. A third metagenome was generated from archived alpine cryoconite for comparison with laboratory-sequenced metagenome data.

For the *Svalbard* metagenomics experiment, community DNA was extracted from Foxfonna cryoconite and sequenced using rapid library preparation, with 3514 reads were successfully basecalled. Of these, through analysis using the Kaiju classifier, 2305 reads could be assigned to named taxa, of which 2265 reads were assigned as bacterial in origin. Consistent with earlier V1-V3 region 16S rRNA gene amplicon sequencing of cryoconite from Foxfonna ice cap(Gokul et al., 2016) the community was dominated by *Proteobacteria* (*Alphaproteobacteria* and *Betaproteobacteria*) followed by *Actinobacteria, Cyanobacteria* and *Bacteroidetes* (FIG 1). Importantly, the most abundant species-level match was to *Phormidesmis priestleyi.* The cyanobacterium *Phormidesmis priestleyi* is frequently detected in Arctic cryoconite and is thought to act as an ecosystem engineer of cryoconite granules,

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

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

The reliance of our *Svalbard* metagenomics experiment described above on internet access and mains electricity represents an important limitation. We therefore sought to conduct a second experiment based in a typical field camp setting. Our *Greenland* metagenome sequencing experiment entailed collection of cryoconite from the margin of the Greenland Ice Sheet and its sequencing. DNA extraction was performed with access to a portable generator, but generator failure required all subsequent steps to be performed on battery power. Quantification powered by a portable power-pack while library preparation was performed without power, and DNA sequencing was performed on laptop batteries. Freezing air temperatures presented challenges for maintaining liquid reagents and optimal MinION temperatures, resolved by storing reagents warmed by body warmth in a down jacket, and operating the MinION in an insulated shield. These limitations constrained the endurance of the sequencing experiment, but of 2372 reads sequenced, 796 reads were assigned to taxonomy and 692 reads matched bacterial taxa (FIG 2). The community revealed was dominated by *Cyanobacteria, Proteobacteria* (*Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria*) and *Actinobacteria,* again consistent with sequence data from

Greenlandic cryoconite(Edwards et al., 2014b; Musilova et al., 2015; Cook et al., 2016). The most abundant species within the profile, *Phormidesmis priestleyi*, represented 21% of the assigned reads. Our *Greenland* metagenome experiment therefore represents the in-field metagenomic sequencing of a microbial community in the resource-limited settings typical of Arctic field camps.

Benchmarking in-field metagenomics

To compare the performance of in-field nanopore sequencing with laboratory based Illumina sequencing of shotgun metagenomes, we re-sequenced cryoconite DNA samples from Rotmoosferner, an alpine glacier, which had previously been characterized using Illumina sequencing(Edwards et al., 2013a). Two experiments were performed. The first experiment was performed in a laboratory setting while the second experiment used reagents and flow cells adapted for ambient shipping and storage.

Comparison of nanopore and Illumina data analysed using Kaiju revealed coherent taxonomic profiles, both internally and to prior analyses (Edwards et al., 2013a) (FIG 3). The Kaiju-profiled communities were dominated by *Proteobacteria* (*Betaproteobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*) with *Burkholderiales* prominent within the community. *Cyanobacteria* were dominated by *Chamaesiphon minutus* matching reads at 7% and 1% of the annotated nanopore and Illumina reads respectively.

Of 6844 nanopore reads, 2767 reads were annotated to taxa, with 2684 matching bacterial taxa while 184283 out of 225144 Illumina contigs were classified by Kaiju, with 182049 matching bacterial taxa. In spite of the frank difference in sequencing effort in the datasets, the percentage of reads classified to bacteria and their taxonomic affiliation is coherent between nanopore and Illumina datasets. Moreover, the performance of Kaiju with nanopore data in our experiments matches the expectations of performance for Illumina data as

351 described by its authors(Menzel et al., 2016). As Kaiju uses protein-level matches to all
352 possible reading frames it likely represents an effective tool for the rapid characterization of
353 the taxonomic composition and implicit functional potential of microbial communities
354 analysed by rapid library in-field nanopore sequencing of metagenomic DNA.

355 An important limitation to all above experiments is the necessity of maintaining a cold-chain
356 for sequencing reagents and nanopore flow cells. This presents logistical challenges for
357 characterizing microbial communities in highly remote locations. We therefore tested an
358 early-release lyophilized field kit provided by Oxford Nanopore Technologies, Ltd. based on
359 the same sequencing chemistry, performing sequencing of the Rotmoosferner samples in a
360 non-laboratory environment using R9.4 flow cells shipped at ambient temperatures and stored
361 for five days. Kaiju classified 18694 out of 31538 reads generated in the experiment, of
362 which 18008 were assigned to the bacterial domain (FIG 4). The taxonomic profile of these
363 reads closely matched both "wet" (RAD003) nanopore and Illumina datasets detailed above,
364 with a coherent higher-level taxonomic profile and the cyanobacterium *Chamaesiphon*
365 *minutus* representing 8% of assigned reads. The performance of lyophilized field kit and
366 ambient flow cells in a non-laboratory environment described above raises the prospect of
367 metagenomic characterization of microbial communities in highly resource limited, remote
368 environments.

369 In field 16S rRNA gene sequencing

370 Amplicon sequencing of 16S rRNA genes represents a backbone technique in microbial
371 ecology for the culture-independent investigation of microbial diversity (eg(Thompson et al.,
372 2017)). In contrast with short read second-generation high throughput sequencing techniques,
373 full-length 16S rRNA gene sequencing is possible with using nanopore(Li et al., 2016;
374 Kerkhof et al., 2017) in laboratory settings. Therefore, we anticipated MinION based in-field

sequencing of barcoded 16S rRNA genes could permit the semi-quantitative comparison of bacterial communities while in the field.

We therefore implemented barcoded 16S rRNA gene MinION sequencing of bacterial communities derived from Svalbard cryoconite holes which were either open (VB1-3; n=3) to the atmosphere, or covered by a layer of snow and superimposed ice (VB4-6; n=3), testing the hypothesis that the presence or absence of snow/ice cover incurs changes in the bacterial community structure.

We amplified and co-sequenced a blank extraction control and a blank PCR control to detect the potential impacts of contamination arising from conducting 16S rRNA gene PCR (Salter et al., 2014), as preparing amplifications in a field laboratory setting may pose additional risks of contamination. A limitation of our experiment was the number of samples which could be processed in a single run, given the capacity of the eight-well PCR portable PCR cycler used. DNA was extracted and amplified from six samples, but PCR product from one sample (VB4, open) was lost during bead clean-up, resulting in sequencing failure for that sample.

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

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

Using the SINTAX classifier algorithm (Edgar, 2016) trained on a highly curated version of the RDP database, reads were classified within ca. 10 minutes on a laptop computer without the need for internet access. At a confidence level of 0.75, 15643 reads per barcode were assigned to bacterial taxa on average (range: 13017-17593, 1 SD = 1725 reads). All community profiles were strongly dominated by *Cyanobacteria* at the phylum level with prominent contributions from *Proteobacteria* (*Betaproteobacteria* and *Alphaproteobacteria*) and *Bacteroidetes* broadly consistent with prior amplicon studies of Svalbard cryoconite (FIG 5). At both the phylum and family level, cryoconite communities ordinated and clustered clearly according to the open or closed status of the cryoconite hole (principal coordinates and hierarchical cluster analysis of fourth root transformed Bray-Curtis distances, FIG 5). At the family level, significant differences were apparent in the fourth root transformed Bray Curtis distances of taxon relative abundance (PERMANOVA: F=6.5, p(MC)=0.02). Our analyses are consistent with the potential effect of snow/ice cover on total bacterial community structure within cryoconite aggregates and underscore the potential for in-field 16S rRNA gene sequencing to inform investigations while deployed in remote environments.

Benchmarking in-field 16S rRNA gene sequencing

To establish whether in-field 16S rRNA gene sequencing implemented as above offers results coherent with established, second-generation sequencing approaches we implemented our nanopore protocol on DNA samples which had previously been characterized by V1-V3 16S rRNA gene pyrosequencing(Edwards et al., 2014b). Our sample set comprised ten cryoconite samples from three Svalbard glaciers (n=2 each) and two alpine glaciers (n=2 each). Negative extraction and PCR controls as above produced one and four reads respectively. From 24000 demultiplexed reads per sample, on average 20882 reads were assigned to bacterial taxonomy (range: 9021-24000, 1 SD =6518). *Cyanobacteria* with prominent contributions from *Proteobacteria* (*Betaproteobacteria* and *Alphaproteobacteria*) and *Bacteroidetes* dominated

the phylum level taxonomic distributions of the community. At the phylum level, significant differences were observed between Arctic and alpine cryoconite communities (PERMANOVA; F=2.8866, p=0.02) with discrete ordination apparent (principal coordinates analysis of Bray-Curtis distances, FIG 6). At the improved taxonomic resolution of family level assignments, the effect of parent glacier is highly significant (PERMANOVA; F=5.032, p=0.001). In this regard, the outcomes of nanopore-based 16S rRNA gene sequencing are highly coherent with prior pyrosequencing analyses(Edwards et al., 2014b).

To further compare the assignment of taxa following nanopore-based 16S rRNA gene sequencing with the pyrosequencing dataset of the same samples, we correlated the log relative abundance of dominant taxonomic groups. Significant, strongly positive Pearson correlations between the log relative abundances of key taxonomic groups were apparent between nanopore and pyrosequencing data: *Acidobacteria* (r=0.82, p=0.004), *Alphaproteobacteria* (r=0.81, p=0.004), *Betaproteobacteria* (r=0.90, p<0.001), *Bacteroidetes* (r=0.70, p=0.02), and *Firmicutes* (r=0.90, p<0.001). The ratios of *Alphaproteobacteria:Betaproteobacteria*, previously identified as a discriminator between Arctic and alpine cryoconite (Cook et al., 2015) were highly correlated between nanopore and pyrosequencing datasets (r=0.90, p<0.001).

Our in-field 16S rRNA gene analyses are limited by two key caveats. Firstly, they are performed on uncorrected nanopore 1D reads and are therefore subject to a relatively high error rate. In this study we anticipated the availability of full-length 16S rRNA gene reads and hence stronger phylogenetic signal relative to the random error profile of nanopore-induced error, coupled with the aggregation of reads at higher taxonomic levels would amortize the effect of diminished taxonomic resolution. Future work in developing nanopore-optimized bioinformatics pipelines for 16S rRNA gene analyses could improve reliable assignment of reads to discrete operational taxonomic units. Secondly, we were constrained

in the number of samples which may be compared in each run due to the limited number of barcodes or PCRs which could be performed. The high throughput of 16S rRNA gene amplicon sequencing we observed on the MinION platform – tens of thousands of demultiplexed reads within a few hours of sequencing – demonstrates the potential for increased parallelization. In summary our analyses highlight the viability of rapid, in-field sequencing of 16S rRNA genes on highly portable MinION devices for comparison of microbial communities.

Portable microbiome sequencing: implications for microbial ecology

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

Advances in high throughput sequencing capacity are underpinning continual revolution, revealing revelations about the structure and potential function of microbial communities inhabiting niches throughout every conceivable habitat in the biosphere. Just over a decade ago, Curtis (Curtis) wrote of the urgent need for microbial ecologists to "go large" and embark upon the high throughput characterization of microbial diversity. Such data-collection initiatives represent essential pre-requisites to the development of mechanistic and predictive insights. Arguably, the vision set out by Curtis (2006) is being accomplished by initiatives such as the Earth Microbiome Project (Thompson et al., 2017), enhancing our coverage of microbial diversity across the planet. However, cataloguing microbial diversity has hitherto been contingent on laboratory-based, high-throughput sequencing platforms, replete with high levels of laboratory infrastructure. This has been at the cost of agility in

472 characterizing and comparing microbial diversity. In-field sequencing using portable MinION
473 sequencers and laptop-based bioinformatics approaches described herein offers the
474 opportunity to regain agility in the characterization of Earth's microbiomes by supporting
475 distributed, at-source DNA sequencing. Loman and Gardy (2017) advocate the merger of
476 human, animal and environmental genomic surveillance through scalable, portable DNA
477 sequencing for digital epidemiology with the goal of achieving a "sequencing singularity".
478 Such a sequencing singularity offers benefits for navigating Earth's microbial diversity in the
479 broader sense also. We contend it is now time to "go small".

480 "Going small" has three conspicuous advantages. Firstly, microbial processes sense and
481 amplify the impacts of environmental change(Vincent, 2010). Understanding the genomic
482 basis of such processes therefore represents a research priority. Parallels may be drawn with
483 the promise of portable sequencing enabled disease diagnosis and surveillance (Gardy and
484 Loman, 2017). For example, within the context of the global cryosphere, rapid warming is
485 creating hitherto unseen opportunities and challenges in the study of microbial diversity.
486 Presently, we lack DNA datasets for over 99.5%(Edwards, 2015; Edwards and Cook, 2015)
487 of Earth's glaciers, which represent Earth's largest freshwater ecosystem (Edwards et al.,
488 2014a) and yet are highly endangered by climatic warming(Hotaling et al., 2017; Milner et
489 al., 2017). Rapid, on-site investigation of microbial interactions with climate warming has the
490 potential to develop baseline data for fragile ecosystems such as glaciers, and to gather
491 observational data as a prelude to hypothesis testing.

492 Secondly, accessing microbial diversity in remote locations, rather than merely its collection
493 and transfer for analysis, also reduces logistical risks. This takes many forms. Sequencing on-
494 site precludes the risk of post-collection changes in community structure, sample degradation
495 or loss in transit *e.g.*(Choo et al., 2015; Hodson et al., 2017). Moreover, investigators gain the
496 flexibility to adjust their sampling campaigns, maximising the value of field campaigns.

Furthermore, retrieval of microbial diversity from inaccessible environments requires sophisticated engineering approaches to assure the integrity of recovered samples. Examples include clean subglacial lake access (Siegert et al., 2012; Christner et al., 2014). In-field sequencing offers the possibility of monitoring the recovery of high quality samples, the detection and prevention of contamination, and the optimization of sample capture events within a single field campaign.

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

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

Conclusions

Here we report the use of portable nanopore sequencing to characterize and compare microbial communities while in the field. Our approaches robustly characterize the

taxonomic composition of glacial microbial communities using shotgun metagenomics, and permit their comparison by 16S rRNA gene amplicon sequencing. The experiments reported show the versatility of nanopore sequencing approaches for microbiome analyses in a range of field settings, and the coherence of data produced with established approaches for investigating microbial diversity. Continued development in the field of nanopore sequencing, for example the use of lyophilized reagents and ambient-temperature stable nanopore flow cells for sequencing with on-laptop analytic approaches raises the prospect of highly agile characterization of Earth's microbiomes at source.

Acknowledgements

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

Author contribution

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

Conflicting interest statement

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

547

548

List of Figures

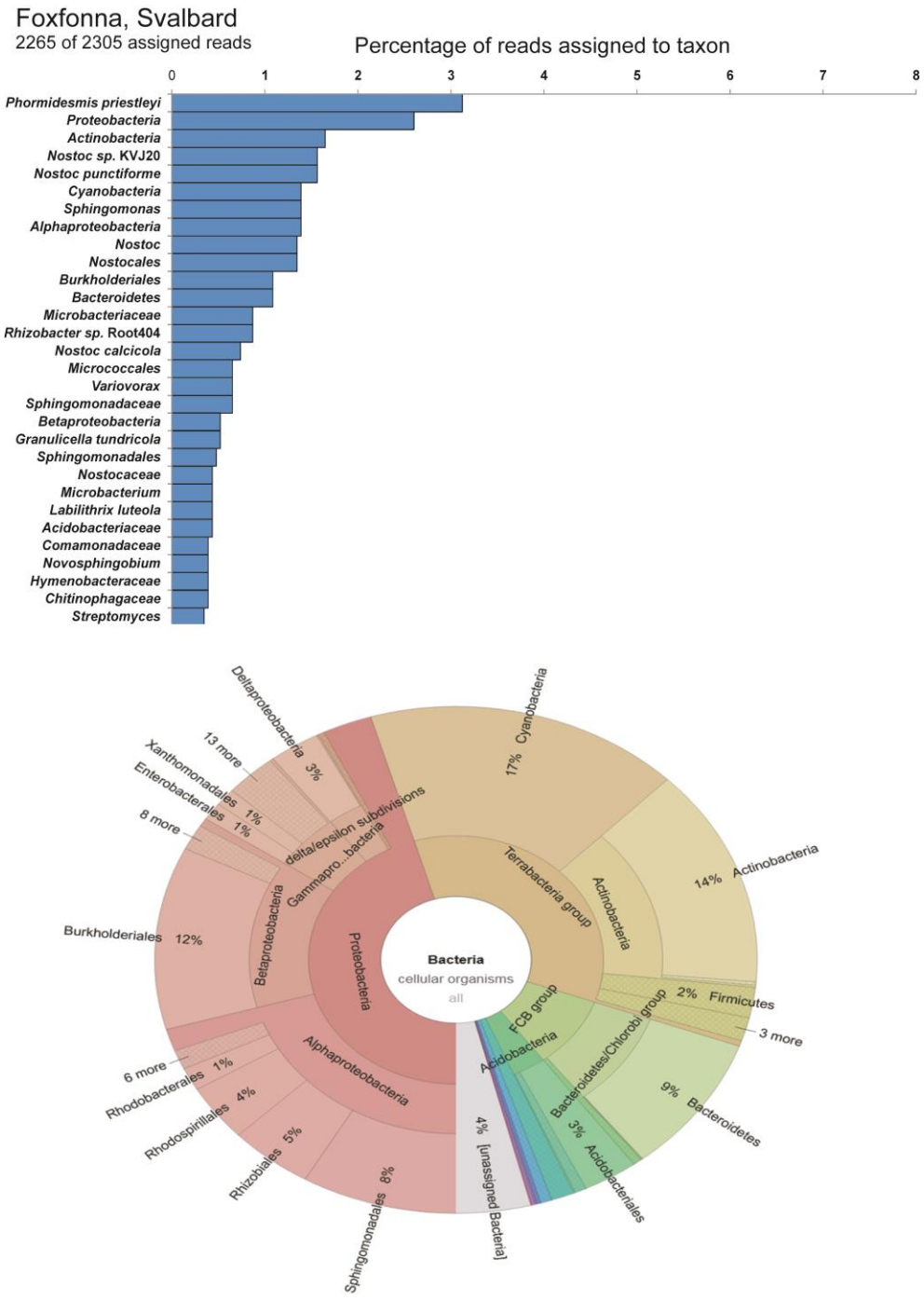
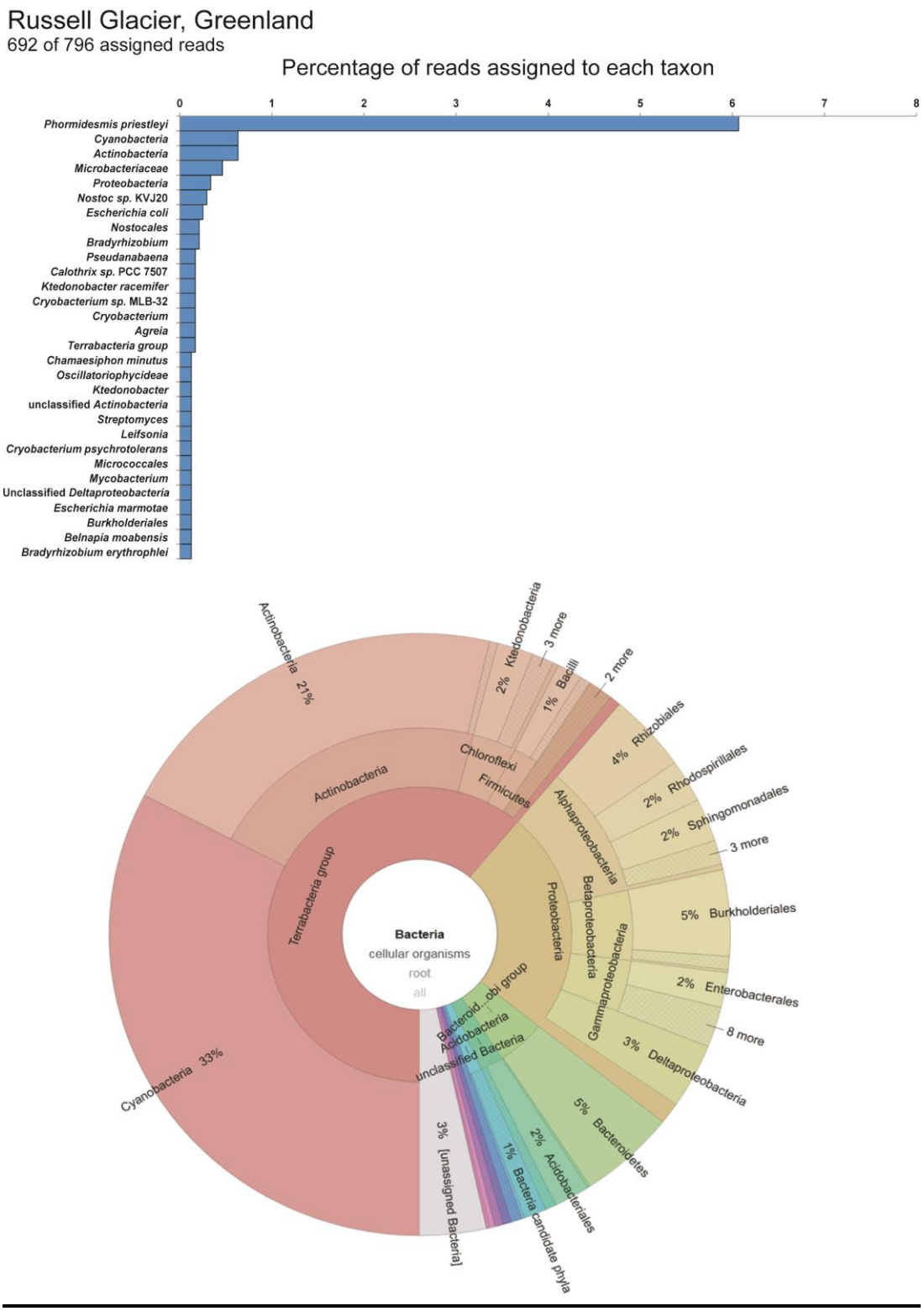
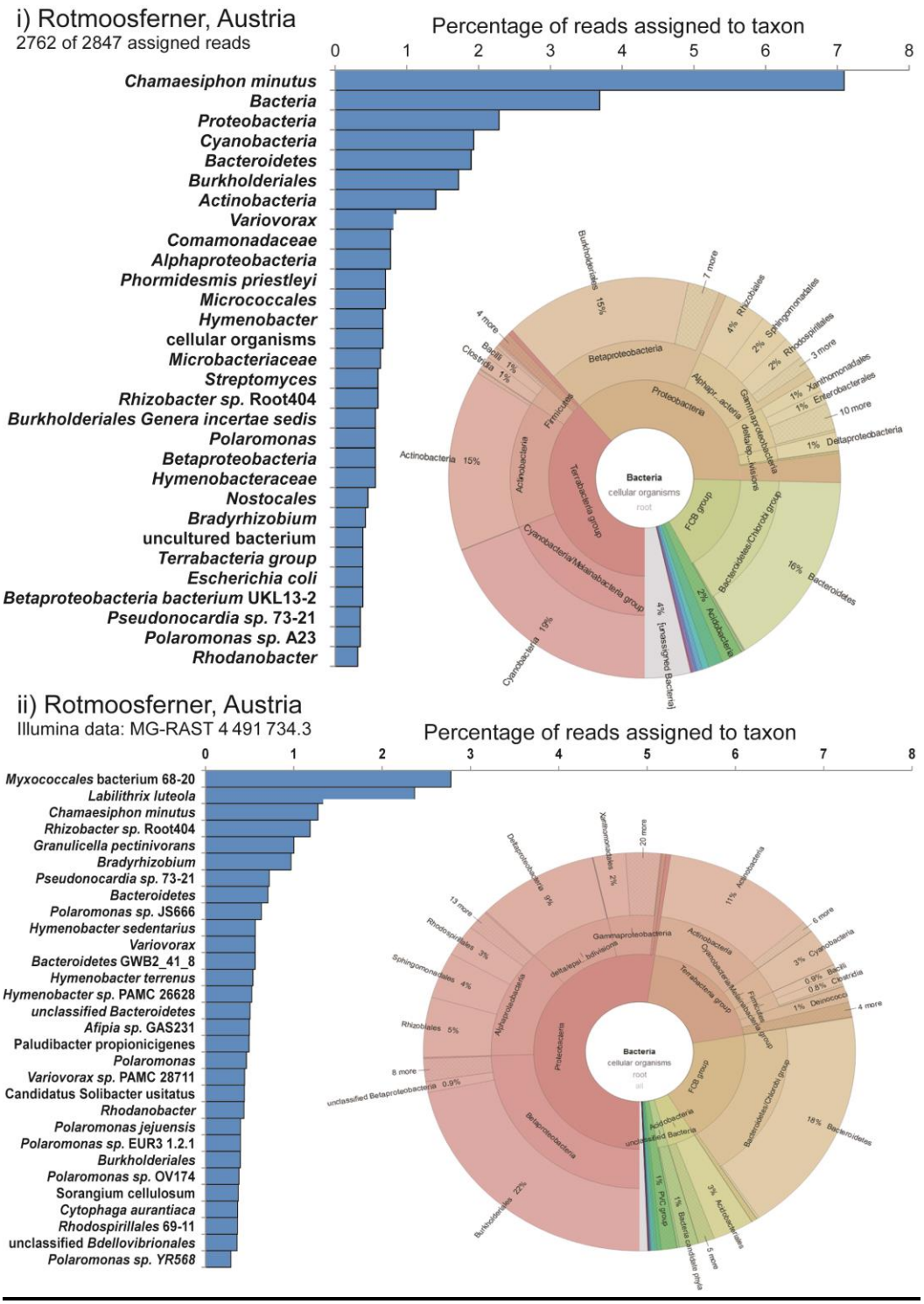


Figure 1: In field metagenome sequencing reveals taxonomic distribution of Svalbard cryoconite microbiota. The thirty top represented taxa and a krona plot of bacterial level diversity in nanopore-sequenced cryoconite metagenome are shown.



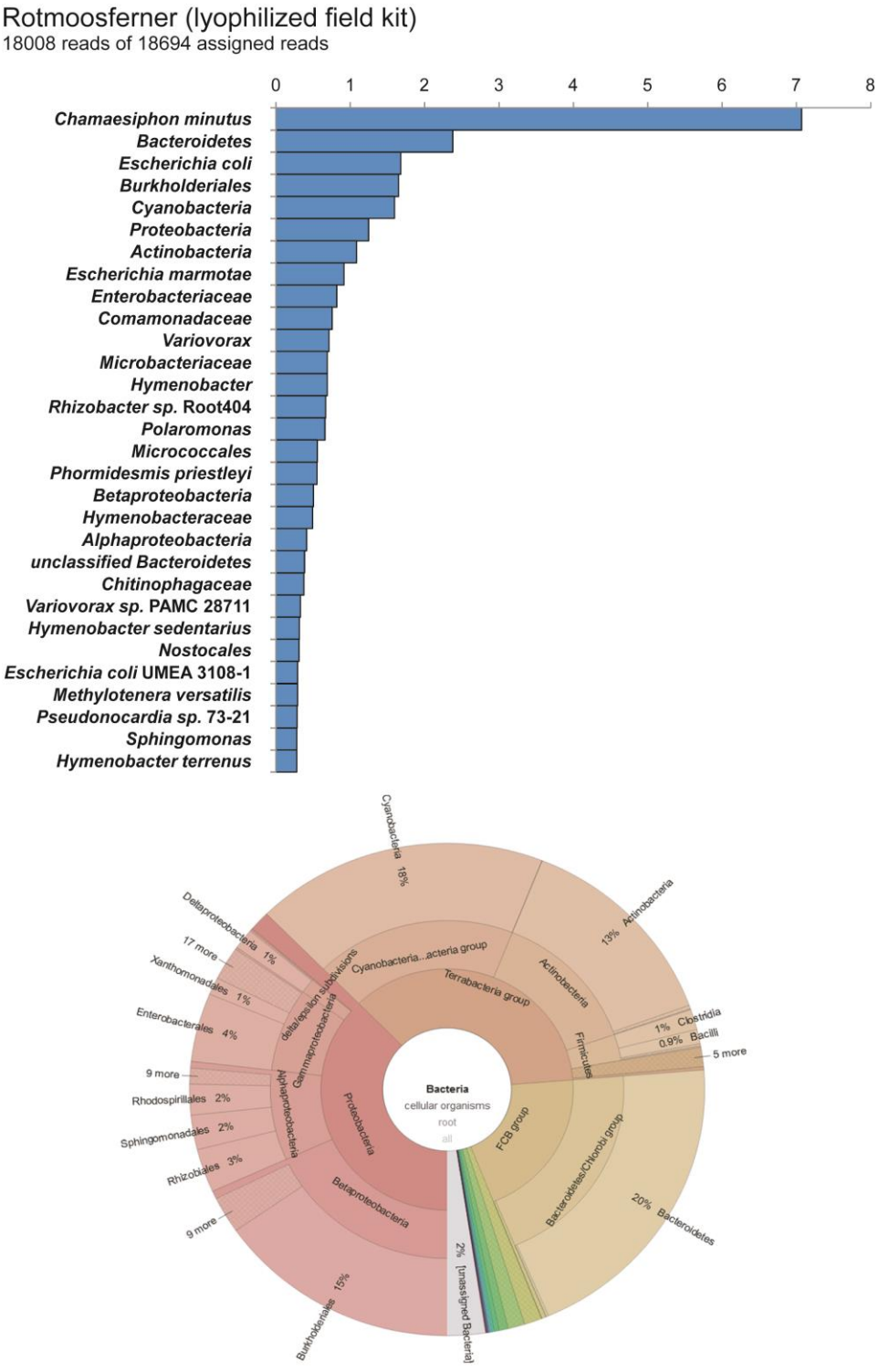
554

555 **Figure 2:** In field metagenome sequencing reveals taxonomic distribution of Greenland
556 cryoconite microbiota. The thirty top represented taxa and a krona plot of bacterial level
557 diversity in nanopore-sequenced cryoconite metagenome are shown.



558

559 **Figure 3:** Benchmarking nanopore metagenomics sequencing by comparison with
560 laboratory-generated data. Krona plot of bacterial level diversity in (i) a nanopore-sequenced
561 cryoconite metagenome from an alpine glacier, Rotmoosferner compared with (ii) Illumina-
562 sequenced metagenome from the same samples (Edwards et al. 2013).



563

564 **Figure 4:** Krona plot of bacterial level diversity in a nanopore-sequenced cryoconite
565 metagenome from an alpine glacier, Rotmoosferner, sequenced using freeze-dried library
566 preparation reagents and ambient-temperature stored R9.4 nanopore flow cells.

567

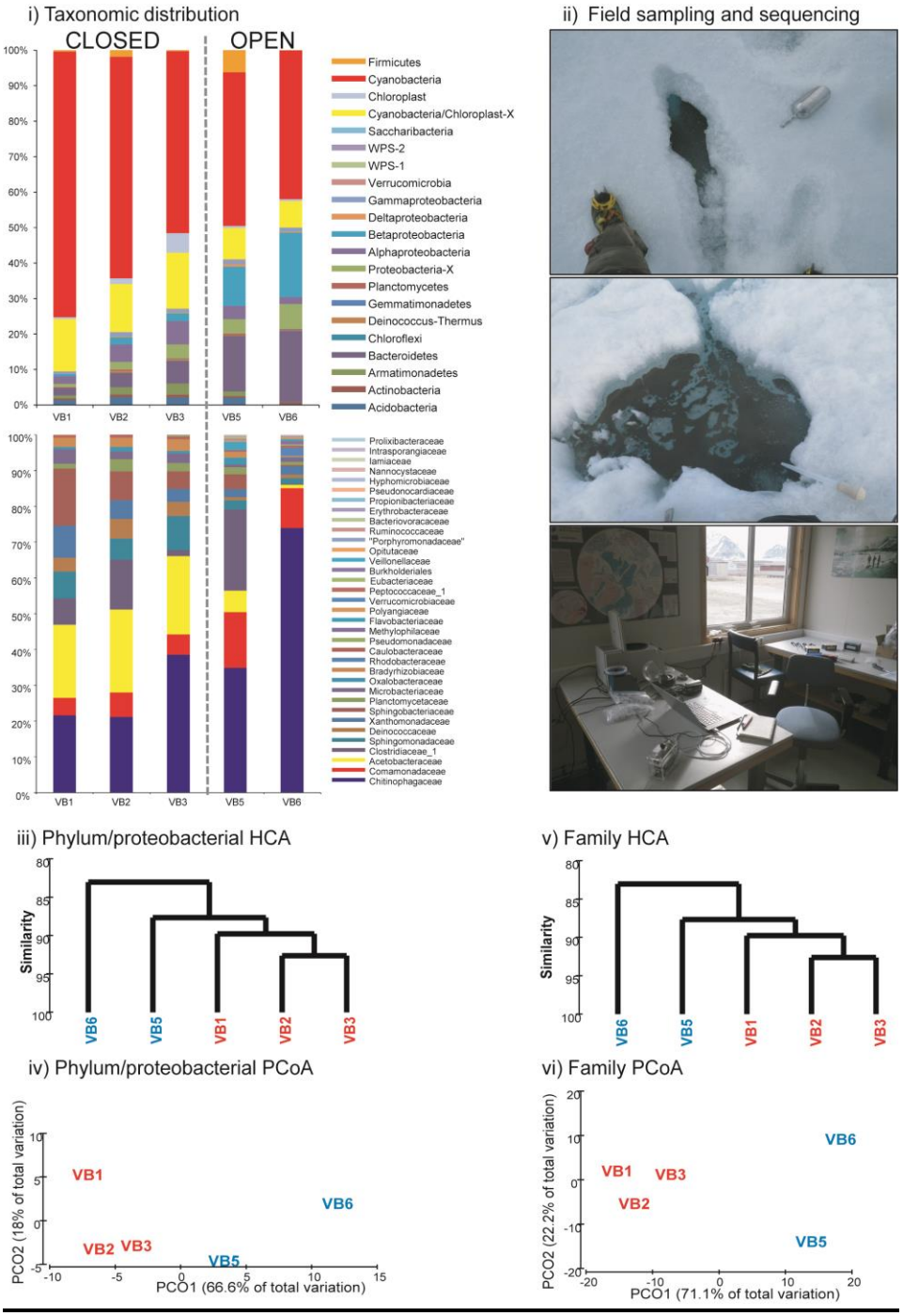
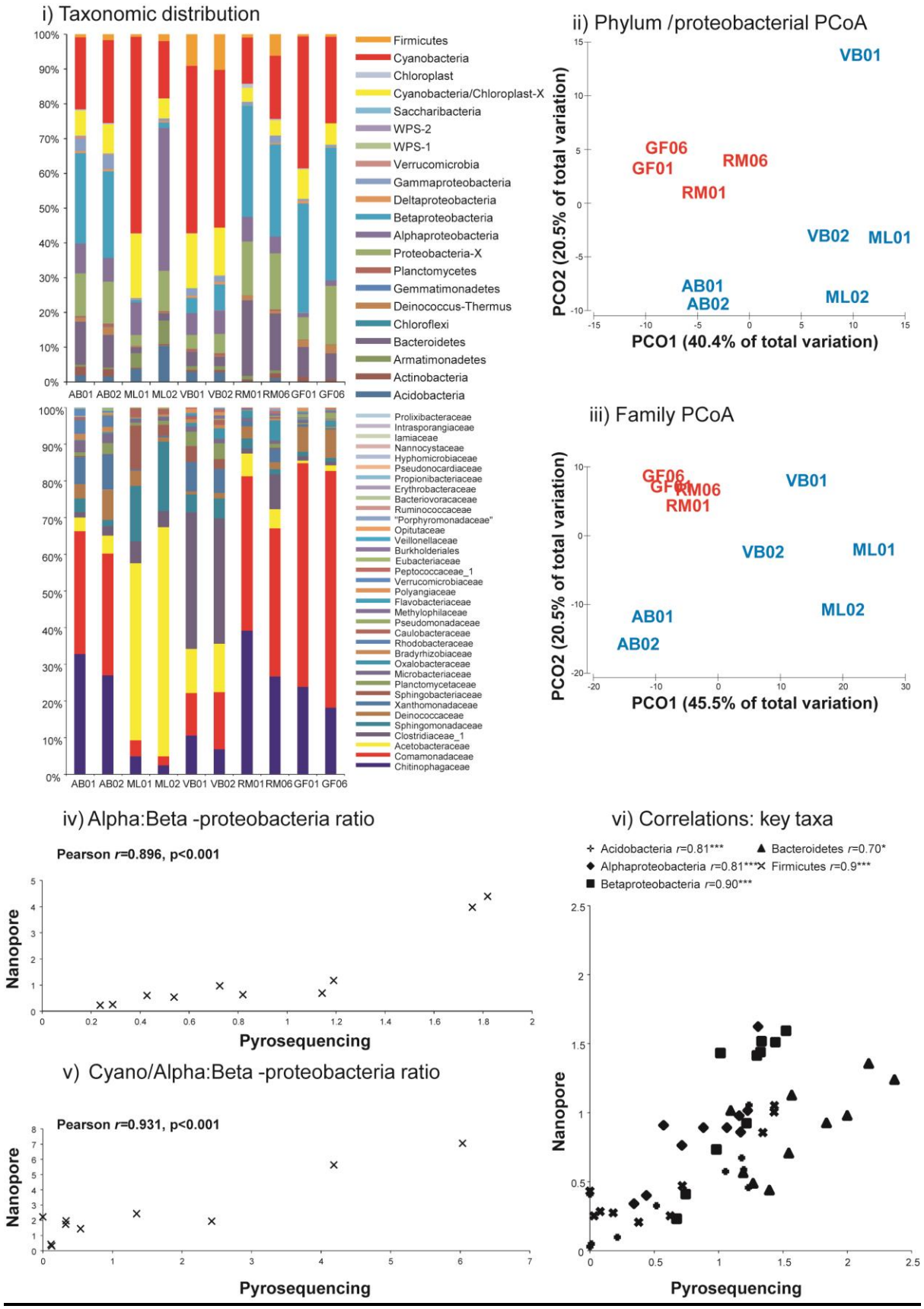


Figure 5: In-field 16S rRNA gene amplicon sequencing of cryoconite bacterial communities showing (i) phylum and family level taxonomic distribution of (ii) open and closed cryoconite holes and multivariate discrimination according to (iii) phylum (iv) and family level taxon distributions following hierarchical cluster analysis and principal coordinates analysis of fourth-root transformed Bray-Curtis distances of phylotype relative abundances.



574

575

Figure 6: Benchmarking nanopore 16S rRNA gene amplicon sequencing of cryoconite bacterial communities by comparison with laboratory-generated data. (i) Phylum and family level taxonomic distribution of cryoconite bacterial communities and (ii) phylum and (iii) family level taxon distributions used for principal coordinates analysis of fourth-root transformed Bray-Curtis distances of phylotype relative abundances discriminate between Arctic and alpine cryoconite communities. Correlation of log relative abundances between (iv) the ratio of *Alphaproteobacteria* and *Betaproteobacteria* (v) *Cyanobacteria to Alphaproteobacteria:Betaproteobacteria* and (vi) key taxonomic groups revealed using nanopore and pyro- sequencing of 16S rRNA genes.

REFERENCES

Anesio, A.M., Hodson, A.J., Fritz, A., Psenner, R., and Sattler, B. (2009). High microbial activity on glaciers: importance to the global carbon cycle. *Global Change Biology* 15(4), 955-960. doi: 10.1111/j.1365-2486.2008.01758.x.

Anesio, A.M., Sattler, B., Foreman, C., Telling, J., Hodson, A., Tranter, M., et al. (2010). Carbon fluxes through bacterial communities on glacier surfaces. *Annals of Glaciology* 51(56), 32-40.

Brown, B.L., Watson, M., Minot, S.S., Rivera, M.C., and Franklin, R.B. (2017). MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach. *GigaScience* 6(3), 1-10.

Calus, S.T., Ijaz, U.Z., and Pinto, A.J. (2018). NanoAmpli-Seq: A workflow for amplicon sequencing from mixed microbial communities on the nanopore sequencing platform. *bioRxiv*. doi: 10.1101/244517.

Cameron, K.A., Hodson, A.J., and Osborn, A.M. (2012). Structure and diversity of bacterial, eukaryotic and archaeal communities in glacial cryoconite holes from the Arctic and the Antarctic. *FEMS Microbiology Ecology* 82, 254-267. doi: 10.1111/j.1574-6941.2011.01277.x.

603 Carbonero, F., Oakley, B.B., and Purdy, K.J. (2014). Metabolic flexibility as a major
604 predictor of spatial distribution in microbial communities. *PloS One* 9(1)**,** e85105.
605 Castro-Wallace, S.L., Chiu, C.Y., John, K.K., Stahl, S.E., Rubins, K.H., McIntyre, A.B.R., et
606 al. (2017). Nanopore DNA Sequencing and Genome Assembly on the International
607 Space Station. *Scientific Reports* 7(1)**,** 18022. doi: 10.1038/s41598-017-18364-0.
608 Choo, J.M., Leong, L.E.X., and Rogers, G.B. (2015). Sample storage conditions significantly
609 influence faecal microbiome profiles. *Scientific Reports* 5**,** 16350. doi:
610 10.1038/srep16350.
611 Christner, B.C., Priscu, J.C., Achberger, A.M., Barbante, C., Carter, S.P., Christianson, K., et
612 al. (2014). A microbial ecosystem beneath the West Antarctic ice sheet. *Nature*
613 512(7514)**,** 310-313. doi: 10.1038/nature13667

614 Cook, J.M., Edwards, A., Bulling, M., Mur, L.A., Cook, S., Gokul, J.K., et al. (2016).
615 Metabolome-mediated biocryomorphic evolution promotes carbon fixation in
616 Greenlandic cryoconite holes. *Environmental microbiology* 18(12)**,** 4674-4686.
617 Cook, J.M., Edwards, A., Takeuchi, N., and Irvine-Fynn, T.D.L. (2015). Cryoconite: the dark
618 biological secret of the cryosphere. *Progress in Physical Geography* in press.
619 Curtis, T. (2006). Microbial ecologists: it's time to 'go large'. *Nature Reviews Microbiology*
620 4(7)**,** 488-488.
621 Darwin, C. (1859). *On the Origin of Species.* London: John Murray.
622 Edgar, R. (2016). SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
623 sequences. *bioRxiv*. doi: 10.1101/074161.
624 Edwards, A. (2015). Coming in from the cold: potential microbial threats from the terrestrial
625 cryosphere. *Frontiers in Earth Science* 3. doi: 10.3389/feart.2015.00012.
626 Edwards, A., Anesio, A.M., Rassner, S.M., Sattler, B., Hubbard, B., Perkins, W.T., et al.
627 (2011). Possible interactions between bacterial diversity, microbial activity and
628 supraglacial hydrology of cryoconite holes in Svalbard. *The ISME Journal* 5(1)**,** 150-
629 160. doi: DOI 10.1038/ismej.2010.100.
630 Edwards, A., and Cook, S. (2015). Microbial dynamics in glacier forefield soils show
631 succession is not just skin deep. *Molecular Ecology* 24(5)**,** 963-966. doi:
632 10.1111/mec.13098.
633 Edwards, A., Debbonaire, A.R., Sattler, B., Mur, L.A., and Hodson, A.J. (2016). Extreme
634 metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N.
635 *bioRxiv***,** 073965.

636 Edwards, A., Irvine-Fynn, T., Mitchell, A.C., and Rassner, S.M.E. (2014a). A germ theory
637 for glacial systems? *Wiley Interdisciplinary Reviews: Water* 1(4)**,** 331-340. doi:
638 10.1002/wat2.1029.
639 Edwards, A., Mur, L.A.J., Girdwood, S.E., Anesio, A.M., Stibal, M., Rassner, S.M.E., et al.
640 (2014b). Coupled cryoconite ecosystem structure–function relationships are revealed
641 by comparing bacterial communities in alpine and Arctic glaciers. *FEMS*
642 *Microbiology Ecology* 89(2)**,** 222-237. doi: 10.1111/1574-6941.12283.
643 Edwards, A., Pachebat, J.A., Swain, M., Hegarty, M., Hodson, A., Irvine-Fynn, T.D.L., et al.
644 (2013a). A metagenomic snapshot of taxonomic and functional diversity in an alpine
645 glacier cryoconite ecosystem. *Environmental Research Letters* 8(3)**,** 035003.
646 Edwards, A., Pachebat, J.A., Swain, M., Hegarty, M., Hodson, A.J., Irvine-Fynn, T., et al.
647 (2013b). A metagenomic snapshot of taxonomic and functional diversity in an alpine
648 glacier cryoconite ecosystem. *Environmental Research Letters* 8(3)**,** 035003.
649 Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The Microbial Engines That Drive
650 Earth's Biogeochemical Cycles. *Science* 320(5879)**,** 1034-1039. doi:
651 10.1126/science.1153213.
652 Gardy, J.L., and Loman, N.J. (2017). Towards a genomics-informed, real-time, global
653 pathogen surveillance system. *Nature Reviews Genetics* 19**,** 9. doi:
654 10.1038/nrg.2017.88.
655 Gilbert, J.A., O'Dor, R., King, N., and Vogel, T.M. (2011). The importance of metagenomic
656 surveys to microbial ecology: or why Darwin would have been a metagenomic
657 scientist. *Microbial Informatics and Experimentation* 1(1)**,** 5.
658 Gokul, J.K., Hodson, A.J., Saetnan, E.R., Irvine-Fynn, T.D., Westall, P.J., Detheridge, A.P.,
659 et al. (2016). Taxon interactions control the distributions of cryoconite bacteria
660 colonizing a High Arctic ice cap. *Molecular Ecology* 25(15)**,** 3752-3767.
661 Goordial, J., Altshuler, I., Hindson, K., Chan-Yam, K., Marcolefas, E., and Whyte, L.G.
662 (2017). In Situ Field Sequencing and Life Detection in Remote (79°26′N) Canadian
663 High Arctic Permafrost Ice Wedge Microbial Communities. *Frontiers in*
664 *Microbiology* 8(2594). doi: 10.3389/fmicb.2017.02594.
665 Hodson, A., Nowak, A., Cook, J., Sabacka, M., Wharfe, E., Pearce, D., et al. (2017).
666 Microbes influence the biogeochemical and optical properties of maritime Antarctic
667 snow. *Journal of Geophysical Research: Biogeosciences*.

668 Hotaling, S., Hood, E., and Hamilton, T.L. (2017). Microbial ecology of mountain glacier
669 ecosystems: Biodiversity, ecological connections, and implications of a warming
670 climate. *Environmental Microbiology*.
671 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., et al.
672 (2016). A new view of the tree of life. *Nature Microbiology* 1**,** 16048.
673 Johnson, S.S., Zaikova, E., Goerlitz, D.S., Bai, Y., and Tighe, S.W. (2017). Real-Time DNA
674 Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer.
675 *Journal of Biomolecular Techniques : JBT* 28(1)**,** 2-7. doi: 10.7171/jbt.17-2801-009.
676 Kerkhof, L.J., Dillon, K.P., Häggblom, M.M., and McGuinness, L.R. (2017). Profiling
677 bacterial communities by MinION sequencing of ribosomal operons. *Microbiome*
678 5(1)**,** 116.
679 Klein, D.A. (2015). Partial Formalization: An Approach for Critical Analysis of Definitions
680 and Methods Used in Bulk Extraction-Based Molecular Microbial Ecology. *Open*
681 *Journal of Ecology* 5(08)**,** 400.
682 Langford, H., Hodson, A., Banwart, S., and Bøggild, C. (2010). The microstructure and
683 biogeochemistry of Arctic cryoconite granules. *Annals of Glaciology* 51(56)**,** 87-94.
684 Li, C., Chng, K.R., Boey, E.J.H., Ng, A.H.Q., Wilm, A., and Nagarajan, N. (2016). INC-Seq:
685 accurate single molecule reads using nanopore sequencing. *GigaScience* 5(1)**,** 34.
686 Lutz, S., Anesio, A.M., Edwards, A., and Benning, L.G. (2016). Linking microbial diversity
687 and functionality of arctic glacial surface habitats. *Environmental Microbiology*.
688 Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for
689 metagenomics with Kaiju. *Nature communications* 7.
690 Milner, A.M., Khamis, K., Battin, T.J., Brittain, J.E., Barrand, N.E., Füreder, L., et al. (2017).
691 Glacier shrinkage driving global changes in downstream systems. *Proceedings of the*
692 *National Academy of Sciences* 114(37)**,** 9770-9778.
693 Musilova, M., Tranter, M., Bennett, S.A., Wadham, J.L., and Anesio, A. (2015). Stable
694 microbial community composition on the Greenland Ice Sheet. *Frontiers in*
695 *Microbiology* 6. doi: 10.3389/fmicb.2015.00193.
696 Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., et al. (2016).
697 Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530(7589)**,**
698 228-232.
699 Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., et al.
700 (2013). Insights into the phylogeny and coding potential of microbial dark matter.
701 *Nature* 499(7459)**,** 431-437. doi: 10.1038/nature12352

Rutter, N., Hodson, A., Irvine-Fynn, T., and Solas, M.K. (2011). Hydrology and hydrochemistry of a deglaciating high-Arctic catchment, Svalbard. *Journal of Hydrology* 410(1-2), 39-50. doi: DOI 10.1016/j.jhydrol.2011.09.001.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* 12(1), 87.

Schmidt, T.M., DeLong, E., and Pace, N. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of bacteriology* 173(14), 4371-4378.

Siegert, M.J., Clarke, R.J., Mowlem, M., Ross, N., Hill, C.S., Tait, A., et al. (2012). Clean access, measurement, and sampling of Ellsworth Subglacial Lake: a method for exploring deep Antarctic subglacial lake environments. *Reviews of Geophysics* 50(1).

Stibal, M., Schostag, M., Cameron, K.A., Hansen, L.H., Chandler, D.M., Wadham, J.L., et al. (2015). Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environmental microbiology reports* 7(2), 293-300.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551(7681).

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428(6978), 37-43.

Uetake, J., Tanaka, S., Segawa, T., Takeuchi, N., Nagatsuka, N., Motoyama, H., et al. (2016). Microbial community variation in cryoconite granules on Qaanaaq Glacier, NW Greenland. *FEMS microbiology ecology* 92(9), fiw127.

van der Gast, C.J. (2015). Microbial biogeography: the end of the ubiquitous dispersal hypothesis? *Environmental Microbiology* 17(3), 544-546. doi: 10.1111/1462-2920.12635.

Vincent, W.F. (2010). Microbial ecosystem responses to rapid climate change in the Arctic. *The ISME Journal* 4(9), 1087-1090.

Watson, M., Thomson, M., Risse, J., Talbot, R., Santoyo-Lopez, J., Gharbi, K., et al. (2015). poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics* 31(1), 114-115.

735 Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T., et al. (2016).
736 Challenges in microbial ecology: building predictive understanding of community
737 function and dynamics. *The ISME Journal*. 10 11:2557-2568.

738