bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 °N

Arwyn Edwards[1,2]*, Aliyah R. Debbonaire[1,2], Birgit Sattler[3] Luis A.J. Mur[1] and Andrew J. Hodson[4,5].

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

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

Keywords: MinION/Nanopore/Metagenomics/Arctic/Glacier/Cryoconite

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 a highly portable USB-based nanopore DNA sequencing platform coupled with field-adapted environmental DNA extraction, rapid sequence library generation and off-line analyses of metagenome profiles to characterize the microbiota of a High Arctic glacier. The profile of the microbial communities produced by this approach are coherent with those from conventional amplicon and shotgun metagenome sequencing of glacier environments, and prompt testable hypotheses regarding microbial community structure and function on the basis of data generated and analysed while in the field. Moreover, we highlight the potential for oligonucleotide mixture model based metagenomic profiling of nanopore reads to deliver off-grid metagenomics analyses. Ultimately, in-field metagenomic sequencing potentiated by nanopore devices raises the prospect of considerably enhanced agility in exploring Earth's microbiomes.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Since most microbes are not readily cultured in the laboratory the exploration of microbial genomic and metabolic diversity underpinning our biosphere's function requires the application of sophisticated biomolecular methodologies (Falkowski et al 2008, Gilbert et al 2010). 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).

The typical *modus operandi* for the investigation of microbial biodiversity via DNA sequencing entails the collection, preservation and transfer of environmental matrices to home laboratories, prior to nucleic acid extraction, high throughput sequencing and bioinformatics analyses. These strategies are well-established, offering both unprecedented depth of insight and opportunities for critique. 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.

For investigators of microbial diversity occurring within "extreme" or remote environments, such as the cryosphere, this retrospective view of microbial diversity presents especial challenges. Firstly, the deployment of investigators to the deep field may be limited to brief seasonal time-windows or entail major logistical complexities which may limit sample capture to singular opportunities per season, for example during the clean capture of Antarctic subglacial lake microbial diversity(Christner et al 2014). Secondly, the amplified climatic warming of the Arctic presents anomalous conditions which may present unpredictable opportunities and challenges for field investigators. For example the spatially extensive yet transient surface melting of the Greenland Ice Sheet in 2012(Nghiem et al 2012) was postulated to affect associated microbial communities (Hell et al 2013, Lutz et al 2016b), but contemporary data to support this notion remains limited(Stibal et al 2015a).

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Finally, since Arctic microbes are themselves recognized as both sentinels and amplifiers of environmental change(Vincent 2010), timely insights to microbial processes interacting with the warming Arctic are imperative. Therefore, greater agility than can be provided by retrograde transfer of materials to home laboratories for subsequent high throughput sequencing is desirable to characterise and predict microbial interactions in extreme environments, and in particular those extreme environments reacting to amplified climatic change or necessitating on-site verification of genomic diversity recovery.

In this study, we report the on-site trial of a highly portable USB-powered, nanopore-based DNA sequencing platform, the minION (Oxford Nanopore Technologies, Ltd) to analyse supraglacial metagenomes while deployed in a minimalistic field laboratory on Svalbard. Hitherto, the promise of the minION DNA sequencing platform to provide on-site genomics has been illustrated within the biomedical context by the sequencing of Ebola virus RT-PCR products for characterizing the West African outbreak(Quick et al 2016), however the application of minION sequencers for shotgun metagenomics based profiling of environmental microbial communities is limited. Nevertheless, advances in nanopore sequencing technology make such a strategy increasingly attractive. In particular, the availability of rapid sequencing library preparation kits permit minimal laboratory infrastructure and lower yields of (meta-) genomic DNA than previous workflows. Therefore, our objective was to implement an on-site workflow for nanopore-based shotgun metagenomic sequencing of supraglacial microbiota which would be compatible with typical logistical arrangements for Arctic fieldwork. We opted to generate shotgun metagenome profiles from cryoconite (Cook et al 2015), a granular microbe-mineral aggregate, and a red snow algal bloom (Lutz et al 2016b). Both microbial habitat types are associated with the biological darkening(Irvine-Fynn et al 2012) of glacier surfaces and their increased melting via albedo feedbacks(Edwards et al 2014a), presenting an imperative for understanding their roles in sensing and amplifying(Vincent 2010) environmental change impacts in the Arctic.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

Field Site and sampling

Svalbard is an archipelago of islands in the European High Arctic and a major locus of research activity in polar glaciology and ecology. Moreover, to our knowledge, the town of Longyearbyen at 78˚ 11'N, 15˚ 32' E) is home to the northernmost wi-fi networks with generally reliable high-bandwidth connections and via the University Centre in Svalbard (UNIS) offers access to general molecular biology laboratories to affiliated researchers. Therefore, considering these range of contingencies and resources we opted to conduct this study from Longyearbyen.

Foxfonna is a small ice cap (surface area ca. 4 km^2) in central Svalbard (78˚ 08'N, 16˚ 07' E) which nourishes two small outlet glaciers, namely Rieperbreen and an un-named outlet glacier draining the northern aspect of the Foxfonna ice cap. For a detailed treatment of the Foxfonna ice cap's glaciological and ecological properties are described elsewhere (Gokul et al 2016, Rutter et al 2011).

In the present study, cryoconite debris was collected from three separate cryoconite holes located on the un-named northern outlet glacier on 18 August 2016 (Day of Year [DOY] 231). A cleaned turkey baster was used for the aspiration of debris to 15 mL sterile centrifuge tubes. A red snow algal bloom was located on the northern aspect of the upper Foxfonna ice cap on August 21 2016 (DOY 234) and sampled using a sterile scoop to deposit biomass to sterile 2 L Whirlpak bags. Cryoconite samples were transferred on ice in ambient temperatures of +2-4°C to a -20°C freezer within a maximum time of three hours, and frozen overnight prior to DNA extraction within the minimalistic field lab. Red snow samples were melted at +4°C overnight and sediment biomass concentrated, initially by gravity flow and subsequently brief centrifugation.

Laboratory Infrastructure

The preparation and sequencing of genomic DNA using the minION sequencing platform requires equipment commonly found in the molecular biology laboratory. We therefore selected compact and multi-purpose items essential for the process of DNA extraction and library preparation which were transported in two items of <23 kg checked luggage. A full list of laboratory hardware is provided as **Supplementary Table 1**, however the key items were an IKA Works, Inc MS2 S8 Minishaker, Gilson, Ltd GmCLab Microcentrifuge,

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Invitrogen Qubit 2.0 Fluorimeter, Hybaid Omn-E 96 well PCR cycler and a full set of Gilson pipettors. Chipped ice was not available, therefore for on-ice incubations, a cold-pack frozen at -20°C was immersed in cold tap water and allowed to equilibrate in a Styrofoam cool box. With the exception of nuclease-free water, all reagents for sequencing used required frozen transport, which was attained by use of packaging with cold packs in Styrofoam box which was itself placed inside a plastic coolbox with further cold packs, which proved adequate for 36h at temperatures ≤ 20°C. MinION flow cells require storage at +2-8°C and were therefore transferred in a Styrofoam box with a chilled cold pack. Aerosol-resistant tips, gloves and disposable laboratory coats were used for all sample manipulations, which were performed on a disinfected table surface. The work environment used provided access to a -20°C domestic fridge-freezer, 240 V AC mains electricity and EDUROAM wi-fi.

Adaptation of environmental DNA extraction protocol

For the present study, community genomic DNA was extracted directly from cryoconite debris in three separate extractions from the cryoconite holes sampled. 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 (Cameron et al 2012, Edwards et al 2011, Lutz et al 2016a, Musilova et al 2015, Stibal et al 2015b).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 an IKA Works, Inc MS2 S8 Minishaker which is designed for the agitation of Bioanalyzer DNA chips (Agilent, Ltd) at a user-determined speed. The dimensions of the Bioanalyzer DNA chip are compatible with the horizontal agitation of snugly-fitted PowerBead tubes, and coupled with the user-determined speed provides reproducible bead beating. Power Bead tubes were inserted flat in the Bioanalyzer DNA chip holder and agitated for 10 minutes at 2,400 rpm. 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

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 event of poor DNA yield, DNA concentration was improved by evaporation of Solution C6 in 20 µL aliquots in a 8-well 0.2 mL PCR strip left uncapped in a Hybaid Omn-E 96 well PCR cycler set to 50 °C until a desired volume was approximated.

Rapid Library Preparation and Sequencing

Quantified community genomic DNA extracts were immediately used as an equimolar pool of each of the three cryoconite DNA extracts, or one red snow DNA extract 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). This protocol subjects 200 ng high molecular weight DNA (typically >30 kbp fragment size; in our experience the PowerSoil® protocol routinely produces DNA fragments 20-25 kbp, although this was not assayed in field conditions) in a volume of 7.5 µL to transposase fragmentation and adapter ligation in a ca. 10 minute protocol with 30°C and 75°C incubation steps in a PCR cycler. 1D sequencing was directly performed using R9 flow cells (Oxford Nanopore Technologies, Ltd) loaded conventionally in 150 µL sequencing mix or directly to the sensor array in 75 µL using a spotON channel. A MinION Mk 1 device run using MinKNOW 1.0.5 (Oxford Nanopore Technologies, Ltd) was used for all sequencing.

Comparison of rapid 1D nanopore sequencing profiles with Illumina metagenome data

Hitherto there are no publicly available shotgun metagenomics datasets from Arctic cryoconite. Therefore, for comparison with conventional lab-based metagenomics we opted to re-sequence a pool of two community genomic DNA samples, RF1 and RF6, from alpine cryoconite collected from the surface of Rotmoosferner glacier in the Austrian Alps in September 2010. V1-V3 16S rRNA amplicon pyrosequencing data for these samples are available at available at EBI-SRA (PRJEB5067-ERP004426) and described elsewhere (Edwards et al 2014b), while 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 2013) and is available at MG-RAST (4 491 734.3). Commuity genomic DNA was extracted from 250 mg cryoconite sediment from samples RF1 and RF6 in 2010 using PowerSoil® as described (Edwards et al 2014b) and stored at -20 °C in the interim. An equal volume pool of

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RF1 and RF6 was sequenced in the UK using a minimalistic field lab using the same arrangements and protocols as implemented on Svalbard for cryoconite.

Bioinformatics Analyses: Cloud and Local Basecalling and Standalone Metagenome Annotation

For the cryoconite metagenomes, 1D basecalling was performed remotely via the Oxford Nanopore Technologies, Ltd. Metrichor platform, while 1D basecalling for the snow algal metagenome was performed locally.

Metrichor includes scripts for the cloud-based taxonomic profiling of metagenomic samples. We implemented *What's In My Pot* [WIMP] plus Basecalling for SQK-RAD001 for the CryoFox-1 cryoconite metagenome with default settings. Reads were grouped by their taxonomic assignments and categories with >5 reads in the dataset were identified manually by reference to NCBI Taxonomy ID.

To evaluate the potential for offline metagenomics we investigated both locally and remotely basecalled metagenomes using an user-friendly standalone taxonomic profiling method. Template strand nanopore reads in .fast5 format were converted to fasta using the pore_parallel GUI of poRe (Watson et al 2015) running in R 3.3.1. Multi-fasta files of nanopore reads were then profiled using heptamer oligonucleotide mixture modelling via Taxy 0.9(Meinicke et al 2011) and taxonomic distribution profiles extracted for comparison. Finally, 1D nanopore reads were compared with the alpine cryoconite assembly described previously (Edwards et al 2013) using Taxy.

A Lenovo® ThinkPad® X1 Carbon running Windows 10 with Intel® Core™ i7-5600U Processor, 8GB RAM and a 512GB solid state drive was used for sequencing and all local analyses.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results and Discussion

Microbial communities from cryoconite and red snow were sampled from Foxfonna, returned to the field lab for DNA extraction, 1D rapid sequencing library generation and nanopore sequencing in the field lab, yielding raw data within ca. 36h of sample collection (**Figure 1**). The principal operational challenges we encountered pertained to the stability of wi-fi signal in the field lab for running MinKNOW and Metrichor, and the yield of DNA for rapid library generation. Further optimization of DNA extraction protocols and the development of local basecalling and standalone metagenomics profiling may mitigate such issues in future. Below, we detail the outcomes of the analyses performed.

Field lab extraction of community genomic DNA

The recovery of high quality, inhibitor free nucleic acid from humic-rich environmental samples such as soil, or in this context cryoconite(Takeuchi et al 2001, Takeuchi 2002) typically requires the use of specialized kits, for example the MoBio, Inc. PowerSoil® DNA extraction kit. In our experience, this kit typically provides DNA fragments with a median size >20 kbp from cryoconite and soils, and yields compatible with the requirements of the 1D rapid sequencing kit (data not shown). We therefore adapted this protocol to provide community genomic DNA in field lab conditions as described. DNA was successfully extracted from three Foxfonna cryoconite samples, each weighing ca. 270 mg wet weight and yielding 22.4 ng μL^{-1} (±1.18 ng μL^{-1}) upon fluorometric quantification. Electrophoretic integrity and A_{260}/A_{280} purity assessments were not performed in the field lab. An equimolar pool of the three samples concentrated briefly by evaporation in a PCR cycler yielded ca. 185 ng DNA in 7.5 µL for 1D rapid sequencing library preparation.

A second 1D library was prepared in the field lab using DNA extracted from ca. 250 mg wet weight concentrated red snow biomass, yielding 11 ng μL^{-1} and a total of 83.25 ng DNA in 7.5 µL for 1D rapid sequencing library preparation.

Sequencing process

One R9 flow cell was used for all experiments, with 1,359 active pores remaining following 36 hours in checked luggage for the journey to Svalbard from Aberystwyth in the United Kingdom. Figure 2 outlines the performance of both sequencing runs conducted on Svalbard, cryoconite (hereafter *CryoFox*) and red snow algal biomass (hereafter *SnowSpot*) as well as a subsequent test of archived alpine cryoconite DNA performed in equivalent field lab settings

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

within the UK (hereafter *AlpsRF1-RF6*). For CryoFox, 3517 reads passing quality filters were analysed, both by WIMP and a standalone metagenomics procedure described below. For SnowSpot and Alps RF1-RF6, very few reads passed quality filters, most likely due to limitations in quantity and quality of input DNA, and thus were analysed by the standalone metagenomics procedure only.

Metagenomic profile of High Arctic cryoconite

WIMP annotation was performed via Metrichor for the CryoFox metagenomic sample. **Figure 3** presents the taxonomic profile generated from WIMP data. Of the 3517 reads passing quality filters, 1533 were assigned to taxonomy at the domain level. Within Figure 3, higher level taxonomic groups are presented, with matches to specific taxa presented within those groups. The reads are dominated by *Bacteria.* Five phyla, including representatives of four proteobacterial classes, are identified by WIMP. Supporting other, amplicon-based studies(Gokul et al 2016) *Archaea* are not detected, but matches to *Eukarya* are present at low abundance, with reads matching ascomycete yeast and unialgal species.

The parallel between the taxonomic distribution annotated by WIMP and 16S rRNA gene amplicon sequencing studies of High Arctic cryoconite is notable. WIMP reveals a community dominated by *Proteobacteria,* with unclassified *Proteobacteria* representing a major proportion of the reads. In common with Sanger based clone libraries, pyrosequencing and semiconductor sequencing of Svalbard cryoconite(Edwards et al 2011, Edwards et al 2014b, Gokul et al 2016) *Alphaproteobacteria* are dominant, with contributions from a range of Sphingomonads consistent with earlier findings from Svalbard cryoconite. Within other heterotrophic bacterial taxa detected, the prominence of *Actinobacteria,* including Micrococci, matches 16S rRNA gene amplicon semiconductor sequencing datasets recently describing a bacterial community prevalent in these groups of *Actinobacteria* in cryoconite colonizing the upper Foxfonna ice cap(Gokul et al 2016). A range of *Betaproteobacteria* taxa are also detected by WIMP; these include groups common in Svalbard glacier surface environments(Rassner et al 2016), and in particular specific assignments to *Polaromonas,* a taxon abundant in melting supraglacial habitats on Svalbard(Hell et al 2013). A preponderance of *Alphaproteobacteria* over *Betaproteobacteria* is consistent with trends observed in 16S rRNA gene amplicon pyrosequencing for Arctic cryoconite versus alpine cryoconite(Cook et al 2015, Edwards et al 2014b).

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cryoconite is characterized as a microbe-mineral aggregate formed by the action of phototrophic bacteria, typically filamentous cyanobacteria(Cook et al 2015). Hitherto, within studies of Arctic cryoconite (Chrismas et al 2016, Cook et al 2016, Gokul et al 2016, Hodson et al 2010, Langford et al 2014) the impact of Oscillatoriophycidae as microbial ecosystem engineers is noted, and in particular the prevalence of *Phormidesmis pristleyi* in cryoconite. Therefore, while the abundance of *Cyanobacteria* within the WIMP profile of CryoFox is consistent with other work, the prevalence of *Nostocaceae*, and in particular reads matching *Nostoc punctiforme* PCC73102 is an unexpected result, especially considering the prevalence of Oscillatoriophycidae (specifically *Phormidesmis*) over *Nostocaceae* in a recent study of Foxfonna cryoconite. BLASTn searches of a randomly selected subset of 50 reads from the 237 reads matching *Nostoc punctiforme* PCC73102 verifies the legitimacy of WIMP annotation.

Local analysis of metagenomes: or, *how to make do with the data you have, rather than the data (and wi-fi) you may wish to have*.

While WIMP annotation of the cryoconite metagenome delivered results largely consistent with the known biodiversity of cryoconite microbiota, its reliance on remote, cloud-based annotation and hence internet connectivity may limit its application in environments with poor or no internet connectivity. Similarly, while WIMP utilizes the goodput of a sequencing run, the challenging nature of environmental samples may necessitate the use of limited quantities of DNA or fragmented community DNA, which limits the data available to WIMP. We therefore tested whether meaningful analyses could be performed without cloud-based resources, and with poorly performing libraries. To this end, we trialled Taxy, an oligonucleotide mixture model based programme(Meinicke et al 2011) which is demonstrated to provide effective, read-length independent and time-efficient desktop PC based taxonomic profiles of metagenome compositions(Meinicke et al 2011, Silva et al 2014), requiring a minimum of user expertise. 1D basecalled reads were converted from .fast5 format to .fasta format using poRe and directly analysed using Taxy, providing taxonomic profiles at phylum, class or order levels within ca. 30 seconds of starting the analysis.

Firstly, the Taxy taxonomic profile (**Figure 4**) of Cryofox inferred a community dominated by *Proteobacteria*, principally Betaproteobacterial *Burkholderiales*, consistent with the presence of *Polaromonas* and other Comamonadaceae revealed by the WIMP profile. Moreover, *Actinomycetales* were prevalent, again mirroring the WIMP profile. The

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Alphaproteobacteria were principally represented by *Rickettsiales* which incorporates the *Sphingomondaceae.* However, the *Cyanobacteria*, a major feature of the WIMP profile and other analyses of High Arctic cryoconite, were notably poorly represented by Taxy. The reasons for this are unclear, but may include a limited representation of cryospheric cyanobacterial genomes(Chrismas et al 2015) within the Taxy profile, although reads assigned to *Nostoc punctiforme* PCC73102 by WIMP were recognized as cyanobacterial by Taxy when tested in isolation (data not shown).

A distinctive profile was revealed for the red snow algal metagenome, with broadly even distribution of nanopore reads between taxonomic groups *Betaproteobacteria, Alphaproteobacteria, Deltaproteobacteria, Actinobacteria*, *Spirochaetes* and *Clostridiales.* Notably, the crenarchaeotal class *Thermoprotei* is detected, consistent with the putative presence of *Archaea* in snowpacks (Lutz et al 2015).

Finally, we wished to compare Taxy-analysed nanopore metagenomic profiles with established metagenomic datasets. Currently there are no publicly available High Arctic cryoconite metagenomes as studies have hitherto focused upon alpine glaciers(Edwards et al 2013, Franzetti et al 2016). Of these, one glacier, Rotmoosferner in the Austrian Alps has both Illumina shotgun metagenomic and V1-V3 16S rRNA gene amplicon pyrosequencing datasets (Edwards et al 2013, Edwards et al 2014b)for the same samples. These studies both revealed a community dominated by *Betaproteobacteria* with a limited representation of *Alphaproteobacteria,* and a predominance of *Actinobacteria* and *Bacteroidetes,* with very few *Cyanobacteria.* Taxy profiling was conducted on a 1D nanopore metagenome from the same DNA extracts as used for amplicon sequencing, RF1-RF6, which were also included in the Illumina shotgun metagenome generated previously. The rapid, offline analysis of 1D nanopore reads offered by Taxy presented a community dominated by *Betaproteobacteria,* specifically *Burkholderiales,* followed by *Deltaproteobacteria* and to a lesser extent, *Alphaproteobacteria,* along with *Sphingobacteriales,* and as such was consistent with the community composition identified by amplicon pyrosequencing(Edwards et al 2014b) and Illumina shotgun metagenome sequencing(Edwards et al 2013). Moreover, Taxy analysis of an assembly of the Illumina shotgun metagenome previously profiled using MG-RAST (Edwards et al 2013) revealed a consistent taxonomic profile with the exception of an under-representation of *Bacteroidetes* relative to the MG-RAST profile, although *Sphingobacteriales* was represented in the nanopore dataset itself.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In summary, considering the general, if not absolute, coherence between conventionally generated and analysed deeply sequenced metagenomic datasets and very shallow nanopore metagenome datasets of lesser quality analysed using rapid offline tools, we contend there is potential for the use of 1D nanopore sequencing coupled with oligonucleotide mixture modelling for the rapid profiling of metagenomes in environments with poor internet connectivity. As such, this strategy may be of value for time-critical or on-the-fly analyses of microbial community composition as a prelude to more formal analyses applying conventional shotgun metagenomic tools.

Can in-field nanopore metagenomics offer ecologically relevant insights?

Important rationales for in-field sequencing include the validation of sampling designs or the iteration of experimental protocols in response to opportunities or challenges experienced in the field. While the analyses described herein illustrate the potential technical feasibility of in-field nanopore metagenomics, do the datasets consequently offer any pertinent insights to the microbial ecology of the habitats sequenced?

Firstly, it is important to acknowledge the limitations of the datasets produced by nanopore sequencing at present: they are limited in sequencing output compared to the leading edge of laboratory high throughput sequencing platforms, and produce sequence data with lower accuracy per base. However, at the time of writing there is no sequencing technology available which can deliver datasets representing the full depth of microbial communities as single contig genomes, so handling an incomplete view of a microbial community is not an uncommon challenge in microbial ecology. Also, the analyses of reads generated within these studies is limited to taxonomic profiling, rather than direct inference of functional potential. However, many important advances in understanding microbial community structure and functionality have been achieved with shallow resolution data e.g.(Fierer et al 2012).

In the context of this study, a notable finding is the prominence of *Nostoc punctiforme* PCC73102 affiliated reads within the CryoFox metagenome and the much lower abundance of *Oscillatoriaceae* affiliated reads than might be expected from amplicon sequencing data of cryoconite communities from the study location and other sites across the High Arctic. Two possibilities account for this observation. Firstly, this may be the result of a taxonomic bias, either in incomplete DNA extraction, or the misannotation of reads. While the DNA extraction protocol employed is very similar to that used in other studies revealing the prevalence of *Oscillatoriaceae*, specifically *Phormidesmis pristleyi*, there is a limited number

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of publically available genomes from cryospheric cyanobacteria, with a genome from *Phormidesmis pristleyi* isolated from cryoconite only available recently (Chrismas et al 2016).

Alternatively, the trend observed is valid, offering an insight to cryoconite ecology. *Nostoc* sp. are indeed documented within High Arctic cryoconite (Gokul et al 2016). A notable variation between *Nostoc punctiforme* and *Phormidesmis pristleyi* is the capability of the former to exhibit cellular differentiation in response to changes in environmental conditions. For example, the formation of akinetes resistant to cold and desiccation stresses in response to phosphate limitation and the formation of heterocysts for nitrogen fixation in oxic conditions in response to nitrogen limitation are well documented (Meeks et al 2002). Indeed, rapid recycling of bioavailable phosphate has been quantified in Svalbard cryoconite (Stibal et al 2009), and late in the melting season the availability of snowpack derived nitrogen may be limited, promoting nitrogen fixation within cryoconite (Telling et al 2011). Consequently, the predominance of *Nostoc* within the cryoconite community may be a function of environmental and nutritional stresses incurred late in the melt season. Importantly, the availability of sequence data while in the field could prompt the formulation of experimentalist approaches to test this hypothesis or to modify experimental approaches to account for the apparent prevalence of *Nostoc*.

Prospects for off-grid metagenomics

Here, our motivation has been to trial the utility of 1D nanopore reads generated using a field-compatible DNA extraction protocol and rapid sequencing kit in a minimalistic field lab for the characterization of microbial communities. In doing so the scope for offline analysis of metagenomic datasets has been explored. This raises the prospect that further optimization of DNA extraction protocols and bioinformatics analyses will permit the generation of metagenomic profiles in the field without access to cloud-based bioinformatics resources or mains electricity. As such a primary limiting factor for the portability of metagenomic profiling would be the endurance of the MinION device and its host laptop. A secondary factor may be the availability of reference genomes from extreme environments for the identification of genomic fragments within datasets generated in the field.

Conclusions

In this study we have demonstrated the technical feasibility of conducting nanopore-based metagenomics for the characterization of microbial communities in extreme environments

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using a highly portable field lab, and explored the potential for oligonucleotide mixture modelling approaches to provide insights into community composition using suboptimal datasets and without access to cloud-based bioinformatics resources such as WIMP. Our next objectives are the further trial of off-grid metagenomics approaches and to cross-validate the capabilities of in-field metagenomics strategies.

Acknowledgements

This work was supported by the HEFCW Research Capital Infrastructure funded Extreme Experimental Environments Laboratory EX^{2}EL Project and logistical support from UNIS. ARD is supported via a PhD studentship from the Life Sciences Research Network Wales. AE thanks Polaris Special Projects. Dr Sara ME Rassner is kindly thanked for the extended loan of her laptop computer. Work in the Austrian Alps was supported by the Society for General Microbiology President's Fund for Research Visits (PFRV10/4) and hosted at the Alpine Research Centre, Obergurgl.

We are grateful to Oxford Nanopore Technologies Ltd for inclusion in the MinION Access Programme, and in particular the excellent support provided by Richard Ronan.

Conflict of interest & Ethics

AE was in receipt of a MoBio, Inc. Microbiome Award in 2014 for his work on glacier molecular ecology, but MoBio, Inc. have had no involvement in any aspect of the present study. This study is conducted under the auspices of Research in Svalbard RiS4584 and RiS6348.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Tables

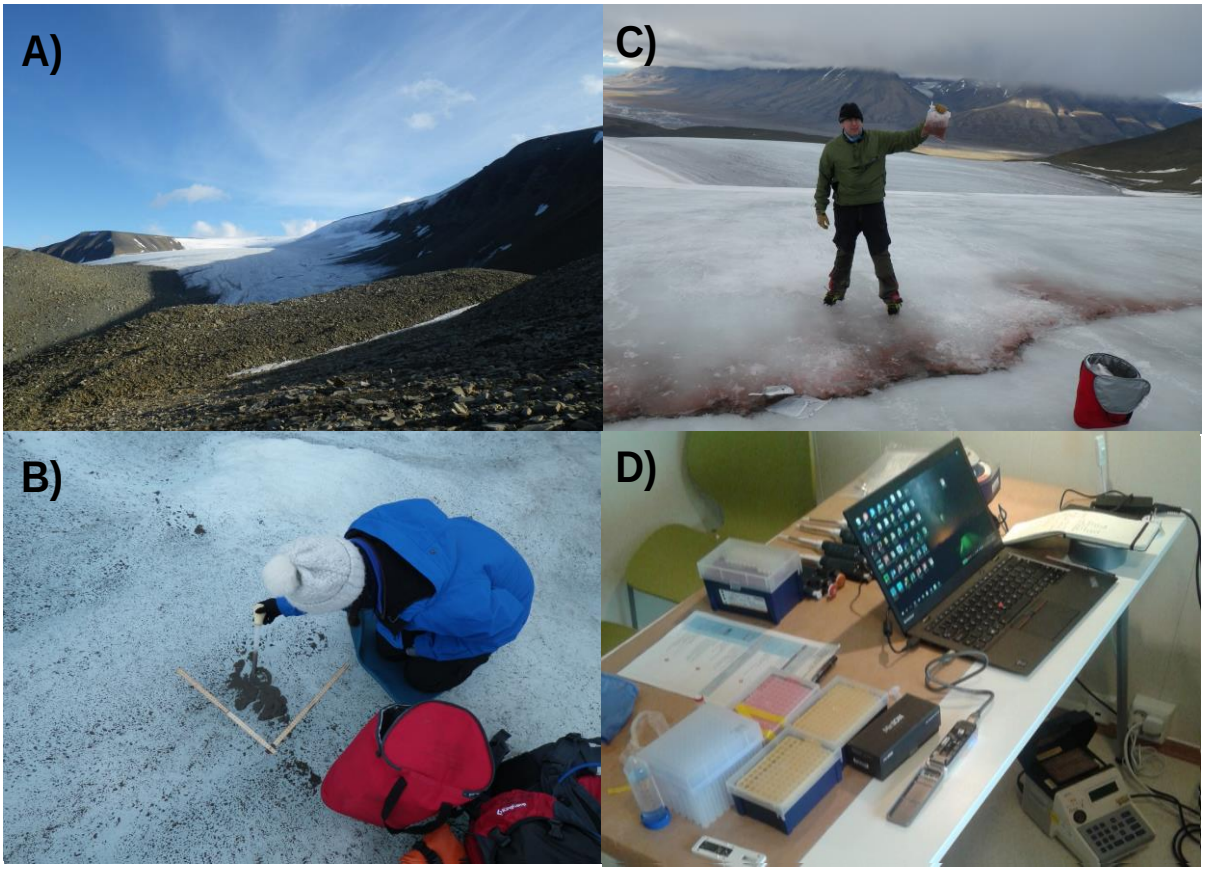

Figure 1: (A): Foxfonna icecap and its un-named North facing outlet glacier (B) Cryoconite sample collection protocol on the North facing outlet glacier of Foxfonna, 18 August 2016 (C) Red snow algal bloom collected on the north-facing aspect of Upper Foxfonna, 21 August 2016. (D) Configuration of field metagenomics lab on a standard domestic tabletop, Visible resources include equipment and materials for DNA extraction, centrifugation, library preparation and MinION sequencing platform.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

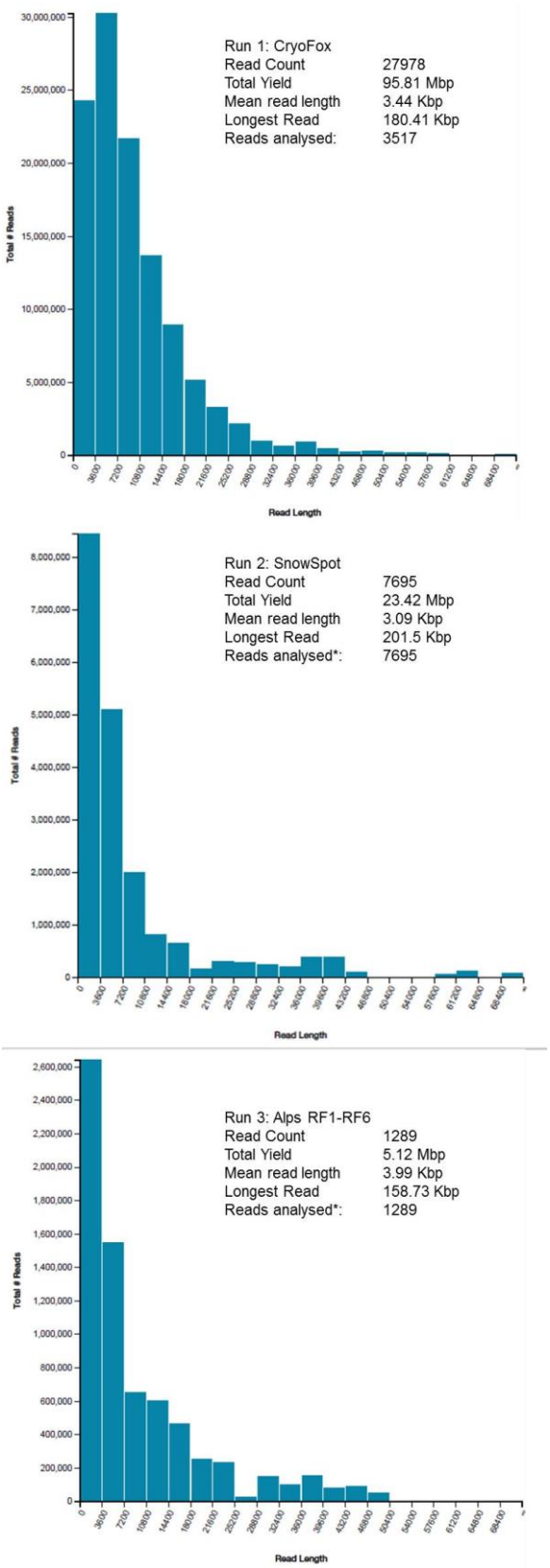

Figure 2: Run statistics and read length histogram for the three 1D rapid sequencing library metagenome runs For Run 2 and Run 3 all reads were analysed using Taxy, while Run 1 reads passing filter were analysed using WIMP and Taxy.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

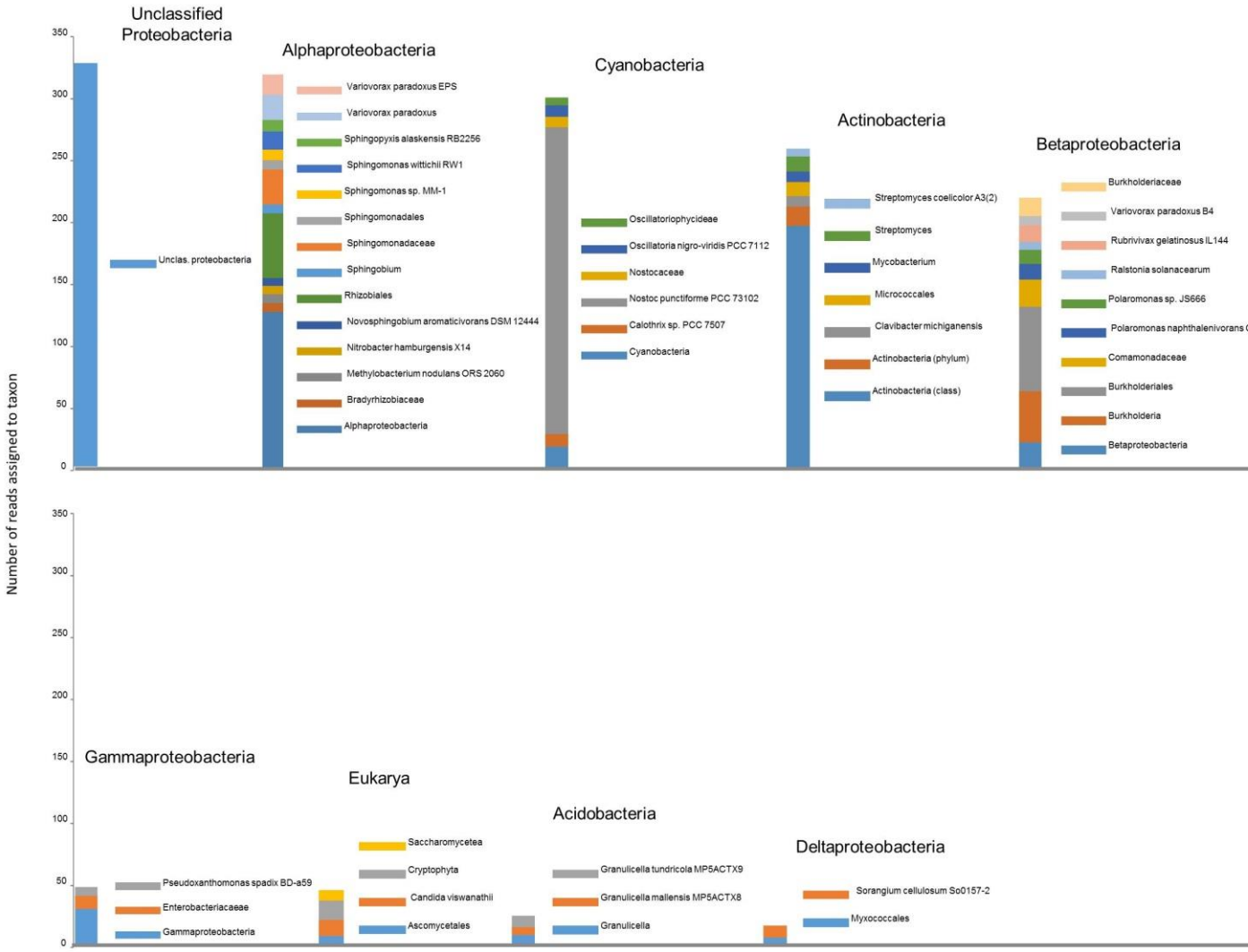

Figure 3 WIMP based annotation of 1533 1D reads assigned to domain level at an abundance of >5 reads per taxon from Cryofox metagenome profile from cryoconite collected 18 August 2016 and sequenced 19-20 August 2016 on Svalbard.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

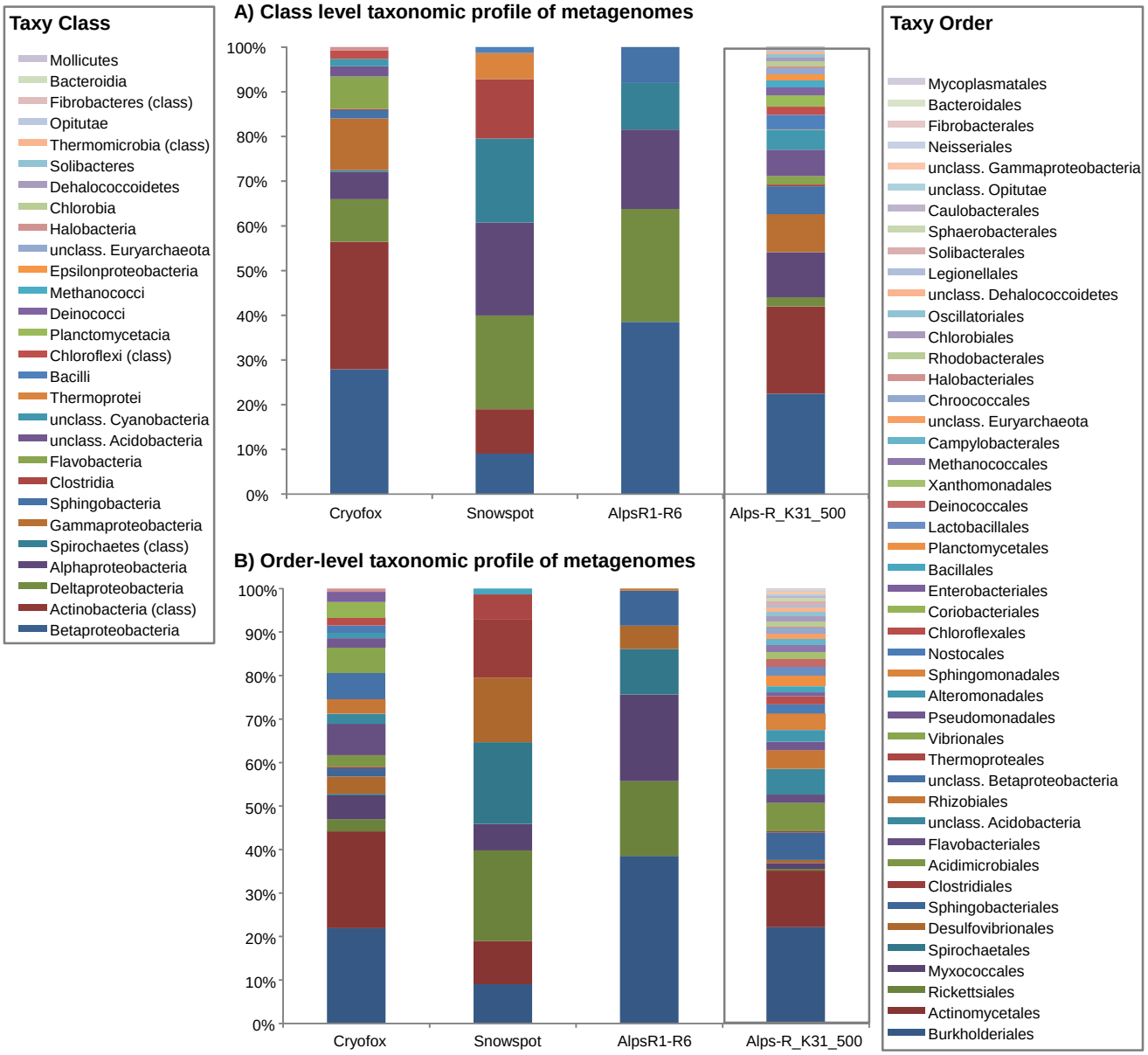

Figure 4: Oligonucleotide mixture model based standalone metagenome profiling using Taxy at class (A) and Order (B) level taxonomic resolutions. Cryofox = metagenome profile from cryoconite collected 18 August 2016 and sequenced 19-20 August 2016 on Svalbard; Snowspot = metagenome profile from red snow bloom collected 21 August 2016 and sequenced 24 August 2016 on Svalbard. AlpsR1-R6 = metagenome profile from alpine cryoconite DNA samples collected from Rotmoosferner and extracted in 2010 (REF) and sequenced on 31 August 2016. AlpsR_K31_500 is a ABySS assembly (K-mer=31, contigs ≥500bp; MG-RAST id 4 491 734.3) of Illumina shotgun metagenomics data from Rotmoosferner cryoconite DNA samples. All Taxy profiles were computed using heptameric profiles.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information

Supplementary Table 1: Field Metagenomics Lab Equipment

Hardware	MinION mk1 device
	Laptop with minKNOW & Metrichor
	4TB Hard Drive
	microcentrifuge
	IKA Vortexor/beadbeater device & cardboard wedge
	Qubit microfluorimeter & protocol
	Pocket scales
	Multiplug
	UK EU electrical adapters
	Hybaid OmnE PCR cycler
	NEB foam rack for Eppendorfs
Pipettors	P2 pipettor
	P10 pipettor
	P20 pipettor
	P200 pipettor x2
	P1000 pipettor x2
Consumables	Gloves, L box
	Gloves, S box
	Duct tape
	parafilm
	LoBind Eppendorf tubes 2x bag
	Mettler-Toledo RT1000 tips for minION loading
	P10 filter tips, 2x box
	P20 filter tips, 2x box
	P200 filter tips, 4x box
	P1000 filter tips, 4x box
	0.2ml 8-strip PCR tubes
	2 litre whirlpaks, sterile
	15 mL centrifuge tubes, sterile
Reagents & transit conditions	
Chill	2x Flow cell, R9
Frozen	SQK-RAD001 rapid sequencing library kit
Frozen	R9 wash kit
Frozen	Blunt TA ligase (NEB) reagent
Frozen	Ampure beads 5 mL
Chill	Qubit hsDNA quantification kit
Ambient	MoBio Powersoil DNA extraction kit
Ambient	Nuclease free water

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Cameron KA, Hodson AJ, Osborn AM (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.

Chrismas NA, Barker G, Anesio AM, Sánchez-Baracaldo P (2016). Genomic mechanisms for cold tolerance and production of exopolysaccharides in the Arctic cyanobacterium Phormidesmis priestleyi BC1401. *BMC genomics* **17:** 533.

Chrismas NAM, Anesio A, Sanchez-Baracaldo P (2015). Multiple adaptations to polar and alpine environments within cyanobacteria: a phylogenomic and Bayesian approach. *Frontiers in Microbiology* **6:** 1070.

Christner BC, Priscu JC, Achberger AM, Barbante C, Carter SP, Christianson K *et al* (2014). A microbial ecosystem beneath the West Antarctic ice sheet. *Nature* **512:** 310-313.

Consortium HMP (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486:** 207-214.

Cook J, Edwards A, Bulling M, Mur L, Cook S, Gokul J *et al* (2016). Metabolome-mediated biocryomorphic evolution promotes carbon fixation in Greenlandic cryoconite holes. *Environmental microbiology* Accepted article: DOI: 10.1111/1462-2920.13349.

Cook JM, Edwards A, Takeuchi N, Irvine-Fynn TDL (2015). Cryoconite: the dark biological secret of the cryosphere. *Progress in Physical Geography* **40**: 66-111

Edwards A, Anesio AM, Rassner SM, Sattler B, Hubbard B, Perkins WT *et al* (2011). Possible interactions between bacterial diversity, microbial activity and supraglacial hydrology of cryoconite holes in Svalbard. *The ISME journal* **5:** 150-160.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Edwards A, Pachebat JA, Swain M, Hegarty M, Hodson A, Irvine-Fynn TDL *et al* (2013). A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. *Environmental Research Letters* **8:** 035003.

Edwards A, Irvine-Fynn T, Mitchell AC, Rassner SME (2014a). A germ theory for glacial systems? *Wiley Interdisciplinary Reviews: Water*. **2**:331-340

Edwards A, Mur LAJ, Girdwood S, Anesio A, Stibal M, Rassner SM *et al* (2014b). Coupled cryoconite ecosystem structure-function relationships are revealed by comparing bacterial communities in Alpine and Arctic glaciers *FEMS Microbiology Ecology* **89:** 222-237.

Falkowski PG, Fenchel T, Delong EF (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320:** 1034-1039.

Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *The ISME journal* **6:** 1007-1017.

Franzetti A, Tagliaferri I, Gandolfi I, Bestetti G, Minora U, Mayer C *et al* (2016). Light-dependent microbial metabolisms drive carbon fluxes on glacier surfaces. *The ISME journal*. doi: 10.1038/ismej.2016.72

Gilbert J, Meyer F, Antonopoulos DA, Balaji P, Brown CT, Brown CT *et al* (2010). Meeting Report: The Terabase Metagenomics Workshop and the Vision of an Earth Microbiome Project. *Standards in Genomic Science* **3:** 243-248.

Gokul JK, Hodson AJ, Saetnan ER, Irvine-Fynn TD, Westall PJ, Detheridge AP *et al* (2016). Taxon interactions control the distributions of cryoconite bacteria colonizing a High Arctic ice cap. *Molecular Ecology*. **25**:3752-67

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hell K, Edwards A, Zarsky J, Podmirseg SM, Girdwood S, Pachebat JA *et al* (2013). The dynamic bacterial communities of a melting High Arctic glacier snowpack. *ISME Journal* **7:** 1814-1826.

Hodson A, Cameron K, Bøggild C, Irvine-Fynn T, Langford H, Pearce D *et al* (2010). The structure, biogeochemistry and formation of cryoconite aggregates upon an Arctic valley glacier; Longyearbreen, Svalbard. *Journal of Glaciology* **56:** 349-362.

Irvine-Fynn TDL, Edwards A, Newton S, Langford H, Rassner SM, Telling J *et al* (2012). Microbial cell budgets of an Arctic glacier surface quantified using flow cytometry. *Environmental microbiology* **14:** 2998-3012.

Klein DA (2015). Partial Formalization: An Approach for Critical Analysis of Definitions and Methods Used in Bulk Extraction-Based Molecular Microbial Ecology. *Open Journal of Ecology* **5:** 400.

Langford HJ, Irvine-Fynn TDL, Edwards A, Banwart SA, Hodson AJ (2014). A spatial investigation of the environmental controls over cryoconite aggregation on Longyearbreen glacier, Svalbard. *Biogeosciences* **11:** 5365-5380.

Lutz S, Anesio AM, Edwards A, Benning LG (2015). Microbial diversity on Icelandic glaciers and ice caps. *Frontiers in microbiology* **6**:307

Lutz S, Anesio AM, Edwards A, Benning LG (2016a). Linking microbial diversity and functionality of arctic glacial surface habitats. *Environmental Microbiology*. doi: 10.1111/1462-2920.13494

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Lutz S, Anesio AM, Raiswell R, Edwards A, Newton RJ, Gill F *et al* (2016b). The biogeography of red snow microbiomes and their role in melting arctic glaciers. *Nature Communications* **7** :11968

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ *et al* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480:** 368-371.

Meeks JC, Campbell EL, Summers ML, Wong FC (2002). Cellular differentiation in the cyanobacterium Nostoc punctiforme. *Archives of Microbiology* **178:** 395-403.

Meinicke P, Aßhauer KP, Lingner T (2011). Mixture models for analysis of the taxonomic composition of metagenomes. *Bioinformatics* **27:** 1618-1624.

Musilova M, Tranter M, Bennett SA, Wadham JL, Anesio A (2015). Stable microbial community composition on the Greenland Ice Sheet. *Frontiers in Microbiology* **6**:193

Nghiem SV, Hall DK, Mote TL, Tedesco M, Albert MR, Keegan K *et al* (2012). The extreme melt across the Greenland ice sheet in 2012. *Geophysical Research Letters* **39**. L20502, doi:10.1029/2012GL053611.

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L *et al* (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530:** 228-232.

Rassner SME, Anesio A, Girdwood SE, Hell K, Gokul J, Whitworth DE *et al* (2016). Can the bacterial community of a High Arctic glacier surface escape viral control? *Frontiers in Microbiology* **7:** 956.

Rutter N, Hodson A, Irvine-Fynn T, Solas MK (2011). Hydrology and hydrochemistry of a deglaciating high-Arctic catchment, Svalbard. *Journal of Hydrology* **410:** 39-50.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Silva GGZ, Cuevas DA, Dutilh BE, Edwards RA (2014). FOCUS: an alignment-free model to identify organisms in metagenomes using non-negative least squares. *PeerJ* **2:** e425.

Stibal M, Anesio AM, Blues CJD, Tranter M (2009). Phosphatase activity and organic phosphorous turnover on a high Arctic glacier. *Biogeosciences* **6:** 913-922.

Stibal M, Gözdereliler E, Cameron KA, Box JE, Stevens IT, Gokul JK *et al* (2015a). Microbial abundance in surface ice on the Greenland Ice Sheet. *Frontiers in Microbiology* **6**.225

Stibal M, Schostag M, Cameron KA, Hansen LH, Chandler DM, Wadham JL *et al* (2015b). Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environmental microbiology reports* **7:** 293-300.

Takeuchi N, Kohshima S, Goto-Azuma K, Koerner R (2001). Biological characteristics of dark colored material (cryoconite) on Canadian Arctic glaciers (Devon and Penny ice caps). *Memoirs of the National Institute of Polar Research* **54:** 495-505.

Takeuchi N (2002). Optical characteristics of cryoconite (surface dust) on glaciers: the relationship between light absorency and the property of organic matter contained in the cryoconite. *Annals of Glaciology* **34:** 409-414.

Telling J, Anesio AM, Tranter M, Irvine-Fynn T, Hodson A, Butler C *et al* (2011). Nitrogen fixation on Arctic glaciers, Svalbard. *Journal of Geophysical Research-Biogeosciences* **116**. 116, G03039, doi:10.1029/2010JG001632.

bioRxiv preprint doi: https://doi.org/10.1101/073965; this version posted September 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37-43.

Vincent WF (2010). Microbial ecosystem responses to rapid climate change in the Arctic. *The ISME journal* **4:** 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:** 114-115.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT *et al* (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal*. doi: 10.1038/ismej.2016.45.