# Illuminating the functional rare biosphere of the Greenland Ice Sheet's Dark Zone

Jarishma K. Gokul[1], Karen A. Cameron[1,2,3], Tristram D.L. Irvine-Fynn[4], Joseph M. Cook[1,5], Alun Hubbard[4], Marek Stibal[6], Matt Hegarty[1], Luis A.J. Mur[1], Arwyn Edwards*[1]

Affiliation: [1]Institute of Biological, Rural and Environmental Sciences, Aberystwyth University, Aberystwyth, UK. [2]Department of Geochemistry, Geological Survey of Denmark and Greenland, 1350 Copenhagen, Denmark. [3]Center for Permafrost, University of Copenhagen, 1350 Copenhagen, Denmark. [4]Centre for Glaciology, Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, UK. [5]Department of Geography, University of Sheffield, Winter Street, Sheffield, UK. [6]Department of Ecology, Faculty of Science, Charles University, Prague, Czechia.

Keywords:

*Corresponding Author: Arwyn Edwards, Institute of Biological, Rural and Environmental Sciences, Cledwyn Building, Aberystwyth University, Aberystwyth, SY23 3DD, UK. aye@aber.ac.uk

Running title: Bacteria in the Dark Zone of the Greenland Ice Sheet

Abstract (222 words)

The Dark Zone of the western Greenland Ice Sheet is the most expansive region of contiguous bare terrestrial ice in the Northern Hemisphere. Microbial processes within the Dark Zone play an important role in driving extensive albedo reduction and amplified melting, yet the composition and function of those consortia have not been fully identified. Here we present the first results from joint 16S rRNA gene and 16S rRNA (cDNA) analysis for the comparison of input (snow), storage (cryoconite), and output (supraglacial stream water) habitats across the Dark Zone over the melt season. Our analysis reveals that all three Dark Zone communities are characterized by a preponderance of rare taxa exhibiting high protein synthesis potential (PSP). Furthermore, taxa with high PSP represent highly connected "bottlenecks" within community structure, consistent with roles as metabolic hubs within their communities. Finally, the detection of low abundance-high PSP taxa affiliated with *Methylobacterium* within snow and stream water indicates a potential role for *Methylobacterium* in the carbon cycle of Greenlandic snowpacks, and importantly, the export of potentially active methylotrophs to the bed of the Greenland Ice Sheet. By comparing the dynamics of bulk and potentially active microbial communities in the Dark Zone of the Greenland Ice Sheet our study provides insight into the mechanisms and impacts of the microbial colonization of this critical region of our melting planet.

## INTRODUCTION

Microbes that colonize snow and ice surfaces live at the critical interface between the atmosphere and cryosphere (Budyko 1969). Their potential to darken glacier surfaces and thereby amplify melt (Cook et al 2017, Lutz et al 2016, Nordenskiöld 1870, Ryan et al 2018, Stibal et al 2017, Takeuchi 2002), has been a long standing question that has recently been identified by the IPCC (AR5) as requiring urgent attention (IPCC 2014). The "Dark Zone" of the Greenland Ice Sheet is a conspicuous band of low albedo bare ice that covers some 10,000 $km^2$ of the western ablating margin of the ice sheet. Surface melt rates of up to eight metres (water equivalent) per year have been observed here, representing a major component to the Greenland Ice Sheet's negative mass-balance and contributor to global sea-level rise (Ryan et al 2016, Van Tricht et al 2016, Wientjes and Oerlemans 2010, Wientjes et al 2011). It is a biologically active surface (Cook et al 2012, Hodson et al 2010) where extensive microbial colonization drives regional surface albedo reduction and enhanced ablation (Ryan et al 2018, Stibal et al 2017, Tedstone et al 2017). Microbial processes associated with Greenland's dark ice surface also contribute to the cycling and hydraulic export of microbial biomass (Cameron et al 2017, Dubnick et al 2017), organic carbon (Bhatia et al 2013a, Musilova et al 2017, Stibal et al 2010), nutrients (Bhatia et al 2013b, Hawkings et al 2016) in significant quantities to downstream englacial, subglacial, and proglacial hydrological networks and ecosystems which ultimately drain to the coast.

Previous studies of microbial diversity within the Dark Zone have focused on supraglacial communities within granular microbe-mineral aggregates termed cryoconite (Cameron et al 2015, Edwards et al 2014a, Musilova et al 2015, Stibal et al 2015) and glacier algae (Lutz et al 2018, Yallop et al 2012). These studies employed transects (e.g.(Edwards et al 2014b, Lutz et al 2018), or used pooled cryoconite material (Musilova et al 2015, Stibal et al 2015), thereby limiting detailed information regarding temporal bacterial community stability. Few

studies have directly addressed the diversity of snowpack bacteria across the region (Cameron et al 2014). Despite the vast scale of this microbial habitat created by seasonal snowmelt (Ryan et al 2019), nothing is known of microbial temporal dynamics within this habitat. Furthermore, although fluvially-exported microbiota from the ice sheet surface may influence downstream biogeochemical processes such as subglacial methane cycling (Dieser et al 2014, Lamarche-Gagnon et al 2019), to our knowledge, the microbial diversity and functional potential of supraglacial meltwater exported from the Dark Zone remains undocumented.

The sequencing of 16S rRNA genes and 16S rRNA (reverse transcribed as cDNA) from co-extracted DNA and rRNA represents a common strategy within microbial ecology. Its application for the discrimination of "total" and "active" bacterial communities has been subject to critique, with limitations in the equivalence of rRNA and "activity" highlighted by Blazewicz et al (Blazewicz et al). With the caveat that ratios between 16S rRNA (cDNA) and 16S rRNA genes are indicative of protein synthesis potential (PSP; (Blazewicz et al 2013), rather than unequivocal quantitative evidence of contemporaneous growth, the technique offers the potential for insights into the responses of taxa to rapidly fluctuating environments. For example, within alpine proglacial streams which experience considerable diurnal fluctuation in temperature and discharge, joint 16S rRNA gene and 16S rRNA (cDNA) sequencing revealed that rare taxa were over-represented in the 16S rRNA (cDNA) population (Wilhelm et al 2014). Within the austere and isolated environs of the Dark Zone in summer, solar radiation, air temperature, melt intensity, and stream discharge all fluctuate with high periodicity, typically diurnally, and hence well within the typical doubling times of supraglacial microbes (Anesio et al 2010, Williamson et al 2018). How the bacterial communities of the Dark Zone respond to these fluctuations remains unknown, and the potential for these rare taxa to disproportionately influence community structure is unknown.

92 In this study, we address these questions by presenting an integrated study of community
93 structure, connectivity, and its functional potential within three principal bacterial habitats
94 within the Dark Zone: snow (input), cryoconite hole (storage) and runoff (output). We
95 evaluate the temporal dynamics of bacterial communities from these three habitats using
96 analysis of both 16S rRNA gene and 16S rRNA (cDNA). This was performed by sampling at
97 weekly intervals, in June and July 2014, to incorporate the transition from early season melt
98 onset to snow-free exposed ice surface.

99

100 ## MATERIALS AND METHODS

101 ### Methods summary

102 Sampling took place on the western ablating margin of the Greenland Ice sheet, within the
103 Dark Zone and adjacent to the Kangerlussuaq (K-) transect S6 automatic weather station
104 (AWS), at 67°05'N, 49°23'W; 1020 m asl (FIGURE 1). Cryoconite, snow and water from
105 supraglacial meltwater streams was collected in triplicate, on seven sampling occasions, at
106 weekly intervals between June 19 and July 31 2014. Independent sites, within a 25m$^2$ area
107 were used for each sampling occasion. A total of 62 samples were collected and chemically
108 preserved using Soil Lifeguard (MO BIO Laboratories). Samples were then frozen as soon as
109 possible. Upon return to the home laboratory, community DNA and RNA was co-extracted
110 from snow and meltwater samples previously concentrated on 0.22 µm Sterivex GP
111 polyethersulfone filters (Millipore, MA, USA) using a modified PowerWater® Sterivex™
112 DNA Kit and from cryoconite using a PowerBiofilm™ RNA Isolation Kit (MO BIO
113 Laboratories) prior to 16S rRNA gene and 16S rRNA (cDNA) quantitative PCR (qPCR) and
114 V3-V4 region MiSeq (Illumina) sequencing. All sequence data are available on EBI-SRA
115 under the study accession number PRJNA318626. The methods employed for sample

archival, nucleic acid extraction, qPCR, sequencing and data processing are detailed in full as **supplementary methods.**

## RESULTS

### Bacterial 16S rRNA gene and 16S rRNA (cDNA) quantification

Quantitative PCR was used to analyse the amplifiable copy number of 16S rRNA genes and 16S rRNA in DNA and cDNA samples (Figure 2). The abundance of the cryoconite bacterial community appeared highly consistent across the sampling period (Figure 2A) with weekly averages of 2.4 - 4.5 × $10^5$ amplifiable copies of the 16S rRNA gene per gram dry weight of cryoconite. The amplifiable copy number of the 16S rRNA pool fluctuated, ranging between a weekly average of 2.3 × $10^7$ (week 4) and 1.6 × $10^9$ (week 2) per gram dry weight of cryoconite. The ratio between 16S rRNA gene and 16S rRNA (cDNA) amplifiable copy number is interpreted as a marker of the overall bacterial PSP. Average cryoconite bacterial PSP showed high (1:88, week 4) to extremely high ratios (1:5000, week 2) throughout the study period. This is indicative of a high level of potential activity relative to biomass within the cryoconite bacterial communities sampled.

For the snow bacterial community (Figure 2B), weekly average amplifiable copies of the 16S rRNA gene increased from 3.0 × $10^4$ copies per litre (water equivalent) in the first week of the study to 7.6 × $10^8$ copies per litre (water equivalent) in the final week. Weekly average amplifiable copies of 16S rRNA (cDNA) also increased from 4.5 × $10^4$ copies per litre (water equivalent) in the first week of the study to 2.8 × $10^7$ copies per litre (water equivalent) in the penultimate week. Weekly average bacterial PSP values for snow were consistently below equivalence, with exception of the first week (1:1.5 ratio). When viewed in the context of the rapid seasonal wastage of the snowpack in the study period, this likely represents the melt

scavenging of bacterial cells incurring physical accumulation of quiescent biomass within the residual snow pack.

Stream water bacterial communities (Figure 2C) exhibited 16S rRNA gene amplifiable copy numbers several orders of magnitude lower in week 1 ($2.8 \times 10^4$ 16S rRNA gene copies per litre) compared to the remainder of the study period ($3.1 - 3.6 \times 10^6$ 16S rRNA gene copies per litre). rRNA copy numbers were at least an order of magnitude higher ($5.5 \times 10^5$ to $3.9 \times 10^5$ 16S rRNA copies per litre) compared to the first week. Stream water bacterial PSP values varied between 0.3 and 2.8 during the study period. In summary, it is likely the stream water bacterial community was an admixture of quiescent and active taxa in transit from different sources (e.g. snowmelt, ice melt, cryoconite and other biofilms) from the ice sheet surface.

Community structure

The total number of reads obtained after sequence processing was 2,673,556 with a maximum of 130,001 reads (GrIScDNAcryo6.3), a minimum of 5 (GrISstream3.1) and a mean of 21,561 reads. Sequences were filtered and rarefied to 943 sequences per sample, resulting in the exclusion of 1 cryoconite DNA sample, 2 stream DNA samples, 3 cryoconite cDNA samples, 3 snow cDNA samples and 6 stream cDNA samples from downstream analysis. Sequences were clustered into 566 operational taxonomic units (OTUs) at 97% sequence similarity. 13.59 % of 16S rRNA gene OTUs and 14.40 % 16S rRNA (cDNA) OTUs were common to all three habitats.

Over 99 % of the sequences in the dataset from cryoconite hole, snow and stream water habitats were successfully assigned to the Greengenes taxonomy using UCLUST. .Non-metric multidimensional scaling of OTUs clearly ordinated both 16S rRNA gene and 16S rRNA (cDNA) profiles of the bacterial communities by habitat type (Figure 3A). These trends were confirmed by PERMANOVA of fourth-root transforms of Bray-Curtis distances

of OTU relative abundance matrices. Highly significant differences were found between all habitat types (Pseudo$F$ = 13.8, $p$ = 0.0001, Supplementary Table 2). Furthermore, highly significant differences are apparent between OTU composition of 16S rRNA gene and 16S rRNA (cDNA) profiles for each of the habitat types (Pseudo$F$ values = 5.3 - 22.5, $p$ = 0.0001). The snow and stream communities revealed in 16S rRNA gene profiles (Pseudo$F$ = 2.6, $p$ = 0.0001 and Pseudo$F$ = 2.4, $p$ = 0.0001 respectively) were temporally dynamic at weekly sampling resolution, with highly significant differences. The 16S rRNA (cDNA) profiles of cryoconite and snow were significant and highly significantly different by week (Pseudo$F$ = 1.7, $p$ = 0.02 and Pseudo$F$ = 3.0, $p$ = 0.0004 respectively), while stream profiles are temporally stable. In contrast, while cryoconite exhibited highly significantly different 16S rRNA gene and 16S rRNA (cDNA) profiles, the 16S rRNA gene profiles were temporally stable. All PERMANOVA results are detailed in Supplementary Table 2.

Trends in taxonomic composition

Pronounced differences between 16S rRNA gene and 16S rRNA (cDNA) taxonomic profiles are apparent for each of the habitats (Figure 3B). Whereas 16S rRNA gene data reveal Actinobacteria, Bacteroidetes and Alphaproteobacteria are the major groups in cryoconite with a modest representation from Cyanobacteria, from the 16S rRNA (cDNA) data Cyanobacteria are the strikingly dominant group throughout the study period (Figure 3). Similarly, while the 16S rRNA gene profiles of snow reveal a transition between Alphaproteobacteria dominated community to a Bacteroidetes dominated community during the study period, Alphaproteobacteria remains the dominant group within the 16S rRNA (cDNA) profile, with an increase in Acidobacteria in the final two weeks of the study. Meanwhile, the discordance between 16S rRNA gene and 16S rRNA (cDNA) profiles are further mirrored in supraglacial streamwater, where Bacteroidetes and Betaproteobacteria were found to be in equitable dominance in the 16S rRNA gene dataset, and

Alphaproteobacteria strongly dominated the 16S rRNA (cDNA) profiles. In summary, the trends in taxonomic composition observed are consistent with discrete bulk and potentially active communities, with phototrophic cyanobacteria active relative to biomass in cryoconite, and the Alphaproteobacteria notably active in snow and stream water.

We considered the potential impact of contamination on the taxa detected in our samples. Negative controls comprising blank DNA extractions were sequenced in parallel with field samples. The controls returned a small number of reads assigned to a total of 10 taxa, and were therefore excluded from the analysis of rarefied data. The most abundant sequence in the control samples was a *Salinibacter* sp. represented by a total of 16 reads (Supplementary Table 3). In contrast, the thirty most represented OTUs from 16S rRNA gene and 16S rRNA (cDNA) data from each habitat match taxa detected orthogonally from the natural environment, with glacial and other cryospheric habitats strongly represented (Supplementary Table 4). The influence of contamination on the present study is therefore considered minimal.

## 16S rRNA gene sequencing

### Trends in potential activity and relative abundance

The strikingly distinctive 16S rRNA gene and 16S rRNA (cDNA) profiles were further investigated to reveal taxonomic groups over- and under-represented in 16S rRNA (cDNA) compared to 16S rRNA genes (Figure 4). All detected phyla/proteobacterial classes with the exception of *Cyanobacteria* were under-represented in 16S rRNA (cDNA) for cryoconite (Figure 4A), whereas Alphaproteobacteria, Acidobacteria, Firmicutes and WPS2 were over-represented in snow (Figure 4B) and Alphaproteobacteria, Verrucomicrobia, OD1 and Firmicutes were over-represented in stream water communities (Figure 4C).

OTUs present at ≤ 1% relative abundance constituted a large percentage of the bulk community within rank abundance curves (Fig 4A-C). Applying the ≤0.1 % of relative threshold commonly used to delimit the "rare" biosphere (Pedrós-Alió 2012), 57 % of cryoconite OTUs, 62 % of snow OTUs and 63 % of stream OTUs would be considered "rare" taxa within the Dark Zone bulk community. Rare taxa would need to exhibit a minimum mean relative abundance of ≥0.005% to be represented in datasets of this size. Rank abundance curves show that 48 % of rare cryoconite-habitat OTUs, 40.45 % of rare snow-habitat OTUs and 42.36 % of rare stream-habitat OTUs exhibit positive protein synthesis potentials (PSP, the ratio between 16S rRNA gene and 16S rRNA [cDNA] relative abundance) over the course of 7 weeks. In each community (Figure 4 A-C), taxon PSP is negatively correlated with mean taxon relative abundance (Spearman correlation; cryoconite: $r$ = -0.63, $p$ < 0.0001, snow: $r$ = -0.65, $p$ < 0.0001, stream $r$ = -0.55, $p$ < 0.0001). The trends exhibited are congruent with the notion that certain rare taxa in surface habitats in Greenland's Dark Zone exhibit disproportionately high protein synthesis potential.

Dynamics of high-PSP OTUs.

To establish the contribution of high PSP OTUs over time, OTUs exhibiting weekly PSP averages ≥1 in the dataset were plotted over time and compared to their relative abundance within the 16S rRNA gene dataset (Figure 5). For the cryoconite community (Figure 5A), 30 of 34 OTUs meeting this criterion were members of Cyanobacteria, with taxa assigned to *Leptolyngbya* representing the majority, including both the highest PSP OTU (*Leptolyngbya*-76) and highest relative abundance OTU (*Leptolyngbya*-3). In the snow community, Alphaproteobacteria represented 22 of 30 taxa with weekly average PSP ≥ 1 (Figure 5A). Notably, *Methylobacterium*-1 is highly abundant in the first week of the study with a corresponding mean PSP of 6.8. However, for the next three weeks, while *Methylobacterium*-1 shows much lower relative abundance, its PSP is strikingly high (ranging 185- to 304-fold).

237 *Methylobacterium*-1 is not detected in snow community after week four. In all, six OTUs
238 assigned to *Methylobacterium* are prominent in the high PSP taxa of the snow community.
239 This is echoed within the stream community. Here, *Methylobacterium*-1 again shows high
240 PSP values, in the range of 49 - 111 between weeks two and seven, however its relative
241 abundance is low, amounting to <2 % of the community overall. Four OTUs assigned to
242 *Methylobacterium* are present among 28 Alphaproteobacteria OTUs, with
243 Sphingomonadaceae taxa well represented. In all, 45 OTUs show weekly average PSP ≥1.
244 Seven *Cyanobacteria* affiliated with *Leptolyngbya* (including the *Leptolyngbya*-3 and
245 *Leptolyngbya*-76 prominent in the cryoconite community) are present with four members of
246 Acidobacteria. The prominence of high PSP taxa in stream water from lineages conspicuous
247 within the snowpack and cryoconite community is consistent with the runoff export of
248 potentially active taxa from surface habitats of the Greenland Ice Sheet's Dark Zone.

249 Keystone species-high PSP rare OTU relationship

250 Taxa that exert a disproportionate influence on the structure of the microbial community,
251 despite low or moderate abundances can be termed keystone species (Power and Mills 1995).
252 High betweenness centrality, measured as the shortest number of paths between any two
253 other OTUs passing through that OTU, is interpreted as a hallmark of a keystone species
254 (Peura et al 2015). Co-occurrence analysis identified sixteen OTUs with betweenness-
255 centrality scores (in the range 7.16 to 0.2; Table 1). All have cumulative positive PSP ratios
256 in at least one habitat over the course of the study, and with the exception of one
257 *Comamonadaceae* OTU with a cumulative mean PSP of 1.98, the remainder show high to
258 very high maxima in their cumulative PSP ratios, in the range 10.5 - 405.8. Again,
259 *Methylobacterium*-1 is represented, with the highest (snow: 405.8) and second highest
260 (stream: 368.2) cumulative mean PSPs. Two other OTUs assigned to *Methylobacterium* show
261 the next highest cumulative mean PSPs in stream and snow. For cryoconite, *Leptolyngbya*

assigned OTUs are prevalent as keystone taxa with high cumulative mean PSP. Of the four *Leptolyngbya* assigned OTUs, *Leptolyngbya*-76, *Leptolyngbya*-106 and *Leptolyngbya*-3 show the highest cryoconite PSPs, but have modest betweenness scores. The considerable overlap between putative keystone species and high PSP taxa, including those present at low abundance, presents the possibility that taxa with high levels of protein synthesis potential are influential in the dynamics of their communities irrespective of their relative abundance. Thirteen of the sixteen OTUs are most closely related to taxa distributed across the global cryosphere (Table 1).

## DISCUSSION

### Snow bacterial communities

Our results provide the first insights into the dynamics of bulk and potentially active communities of decaying snowpacks in the ablating zone of the ice sheet during the transition to bare ice. At the start of the study bacterial PSP is positive (Figure 2), however this rapidly declines for the remainder of the sampling period. In contrast, 16S rRNA gene copy numbers increase by 3-4 orders of magnitude for the remainder of the sampling period. This is likely due to the accumulation of biomass within decaying snow due to physical processes rather than biological growth (Björkman et al 2014). Melting snowpacks are physically and chemically dynamic environments, and it appears only a few lineages are able to maintain their populations in supraglacial snow as it decomposes to slush (Hell et al 2013), with other taxa being washed out. Indeed, 16S rRNA gene and 16S rRNA (cDNA) OTU profiles were highly significantly different over time (Supplementary Table 2).

Here, snowpack 16S rRNA gene copies greatly exceed 16S rRNA copy numbers, indicating the bulk community is likely to be exported as cells with low PSP. For example, the relative

under-representation of Bacteroidetes in 16S rRNA (cDNA) raises the possibility that cellulose-degrading taxa become quiescent when dissociated from sources of complex organic carbon, for example supraglacial phototrophs (Smith et al 2016). It is therefore likely that the abundant groups of bacteria in decaying snow serve as sources of cellular carbon and nutrients rather than viable taxa capable of inoculating downstream habitats. The rare, high PSP *Methylobacterium* sp. OTUs detected represent an exception which will be discussed below.

Although most of the Greenland Ice Sheet is perennially covered with snow, few studies have examined the snowpack microbiology of the Greenland Ice Sheet (Cameron et al 2014). Moreover, the highly isolated setting of field sites coupled with the potential for contamination of low-biomass samples make such studies challenging. By establishing a field camp for the duration of the study, careful handling of samples and the sequencing of negative controls we were able to mitigate these limitations. Negative controls returned very small numbers of reads (Supplementary Table 3). Prominent groups of bacteria in our study were not represented in negative controls with the exception of seven reads matching *Phormidesmis priestleyi*, likely indicating post amplification carry-over of a dominant amplicon type at negligible levels compared to its abundance in field samples.

Cryoconite bacterial communities

Lower copy numbers of 16S rRNA genes were amplified by qPCR compared to previous studies employing 16S rRNA gene qPCR based upon larger, wet-weight samples (Stibal et al 2015). However, the overall trends are consistent between both studies. Considering potential limitations in extraction efficiency and biases inherent in all PCR based analyses, we avoid

309 treatment of qPCR data as absolute quantities of 16S rRNA genes or 16S rRNA in our
310 samples and limit our comparison to trends within the dataset.

311 Exceptionally high ratios of 16S rRNA to rRNA genes were measured in cryoconite (Figure
312 2). Combined with amplicon sequencing data revealing cyanobacteria were overwhelmingly
313 dominant within the 16S rRNA (cDNA) population (Figure 3, Figure 5) we interpret this as
314 evidence of the high PSP of cyanobacteria within the cryoconite granules. Since filamentous
315 cyanobacterial phototrophs such as *Phormidesmis priestleyi* are well known as ecosystem
316 engineers (Cook et al 2015, Edwards et al 2014a) of cryoconite granules through their
317 primary production and granule-building (Langford et al 2010), this is highly plausible.

318 Other work within the same field season at the same site lends support to our findings.
319 Firstly, *Phormidesmis priestleyi* was isolated in culture and genome sequenced (Chrismas et
320 al 2016) and secondly perturbation of cryoconite hole structure and microbial activity
321 revealed *Phormidesmis* sp. employ sensitive photoadaptive mechanisms to optimize carbon
322 sequestration in cryoconite holes (Cook et al 2016). Correspondingly, the prominence of
323 OTUs extremely closely related to *Phormidesmis priestleyi* (Table 1), albeit assigned to
324 *Leptolyngbya* (-3 and -76) within the high PSP (Figure 5) and keystone taxa (Table 1) of
325 cryoconite granules extends insights from previous studies by linking specific *Phormidesmis*
326 lineages with metabolic activities within Arctic cryoconite.

327 Importantly, previous analysis of 16S rRNA genes has resolved a single *Phormidesmis* OTU
328 is cosmopolitan (Segawa et al 2017) within diverse Arctic glacial settings (Cook et al 2016,
329 Edwards et al 2011, Gokul et al 2016, Uetake et al 2016). However, in our study, while one
330 *Phormidesmis* OTU (*Leptolyngbya*-3) is likely to play a structural role, two other lineages
331 (*Leptolyngbya*-76, *Leptolyngbya*-106) have PSP in gross excess to their biomass, indicated by
332 contrasting trends in PSP and 16S rRNA gene relative abundance (Figure 5). The bulk

bacterial community structure of cryoconite granules was stable over the course of the study, consistent with prior studies (Musilova et al 2015). However the *Phormidesmis* dominated 16S rRNA (cDNA) pool of the bacterial community of cryoconite changed over time, with highly significant changes revealed by PERMANOVA (Supplementary Table 2). Therefore, the potential for metabolic and structural niche differentiation among cryoconite *Phormidesmis* merits further investigation.

Supraglacial stream water bacterial communities

Meltwater runoff from the Greenland Ice Sheet surface is thought to be a major contributor to sea level rise (Smith et al 2017). Although this meltwater is an important source of organic carbon and nutrients to downstream ecosystems (Bhatia et al 2013a, Bhatia et al 2013b, Hawkings et al 2016, Musilova et al 2017) and the microbial fluxes in outflows of the Greenland Ice Sheet have been studied (Cameron et al 2017, Dubnick et al 2017), the absence of data on microbial export from the Greenland Ice Sheet surface represents a critical lacuna in our understanding of the Greenland Ice Sheet ecosystem. Within this study, this is addressed by 16S rRNA gene and 16S rRNA (cDNA) qPCR and sequence data which reveal the export of microbial groups prevalent in snow and cryoconite in meltwater from three ephemeral supraglacial streams.

Quantitative PCR reveals the stream water bacterial community in the first week of the study contains approximately equal copy numbers of bacterial 16S rRNA genes and 16S rRNA (cDNA) resulting in a bacterial PSP of 0.97. By the second week, both genes and rRNA have increased their average copy number by two orders of magnitude. Highly significant differences were observed in the community structure of bulk, but not active stream bacterial communities over time (Supplementary Table 2). Only one profile each of the bulk and active communities could be analysed from week one, but both were strongly dominated by

Alphaproteobacteria. Subsequent weeks are marked by a more diverse bulk bacterial community, although *Alphaproteobacteria* were highly dominant in the potentially active bacterial community throughout. Sphingomonadaceae, Acetobacteraceae, and Rickettsiaceae are prevalent in the high PSP *Alphaproteobacteria* detected in stream water, but the highest PSP taxon is *Methylobacterium*-1, which is also a high-betweenness putative keystone species.

The presence of cyanobacterial taxa associated with cryoconite, including *Leptolyngbya-3* and *Leptolyngbya*-76 in stream water indicates the export of potentially active primary producers. It is likely these cyanobacteria originate from biomass sheared from cryoconite granules either present in cryoconite holes, as distributed cryoconite on the ice surface or in stream cryoconite (so-called "hydroconite"; (Hodson et al 2007). It is likely these phototrophs represent sources of highly bioavailable dissolved organic carbon exported from the glacier surface (Musilova et al 2017).

<u>High Protein Synthesis Potential Rare Taxa in the Dark Zone of Greenland</u>

A consistent pattern for all three habitats sampled was that bulk and potentially active communities of snow, cryoconite and stream habitats were highly significantly different (Supplementary Table 2, Figure 3). Furthermore, a small subset of taxa were consistently over-represented in 16S rRNA (cDNA) compared to their corresponding 16S rRNA gene relative abundances, most notably the Cyanobacteria in cryoconite and Alphaproteobacteria in snow and stream habitats. Other taxa were typically under-represented. The majority of taxa present within the surface of the Dark Zone appear to exhibit low protein synthesis potential. This may be due to resource limitation, dormancy or the detection of DNA associated with dead cells (Blazewicz et al 2013). Each of the above scenarios has important ecological implications. Firstly in terms of maintaining a pool of organisms which may

respond to stimuli such as allochthonous resources, and secondly, the maintenance of a "seedbank" of conditionally viable cells (Lennon and Jones 2011), or at the very least, the contribution of carbon and nutrients in otherwise oligotrophic environments in the form of necromass. Differentiation between these scenarios is beyond the scope of 16S rRNA gene analyses (Blazewicz et al 2013) and the potential for "active" taxa to be mis-classified as "dormant" by 16S rRNA (cDNA) / 16S rRNA gene comparison has been described in computational simulations (Steven et al 2017).

Nevertheless, a further notable trend which was consistent across all three habitats was a marked prominence of rare taxa among those with high PSP. Such patterns have been observed in other, comparable contexts, for example within proglacial streams in the European Alps (Wilhelm et al 2014). In those systems, such trends have been considered hallmarks of habitats exhibiting severe fluctuations in their environmental conditions such as temperature, discharge or solar radiation at timescales briefer than the doubling time of the resident community (Lennon and Jones 2011, Wilhelm et al 2014). Here, no overall trends were apparent in terms of the influence of meteorological parameters from week to week, however when monitored continuously (Figure 1) profound oscillations in temperature, incoming short- and long- wave radiation, air temperature, energy flux and melt intensity are apparent at diurnal to sub-weekly periodicity. Considering the sluggish growth of organisms at temperatures at or near freezing (Anesio et al 2010), it is likely that such oscillations create rapidly changing niche spaces at timescales shorter than community doubling times. It is therefore likely that organisms exhibit high PSP relative to their biomass when they are able to maintain activity but not achieve net population growth in the face of unstable fitness.

Consequently, through their metabolic activities, high PSP rare taxa may be disproportionately influential within their communities, fitting the definition of keystone taxa. This is coherent with the corresponding prevalence of high PSP rare taxa among putative

keystone taxa identified by betweenness (Table 1). Gokul et al (Gokul et al) previously identified supraglacial keystone taxa in the cryoconite communities of a High Arctic ice cap, and the present study extends the case that specific rare taxa exert disproportionate influence on the bacterial communities of glacier surfaces through maintaining high levels of metabolic potential. Table 1 reports these taxa typically possess very close relatives (either as environmental sequences or named isolates) in a diverse range of habitats within the global cryosphere, with two implications. Firstly, this lends pragmatic support to their likely authenticity within the communities of the Greenland Ice Sheet Dark Zone, but secondly, the inference is that adaptations resulting in disproportionately high PSP may be common among cosmopolitan species in the polar and alpine regions.

## Implications for biogeochemical cycling

The prominence of OTUs assigned to *Methylobacterium* in the high PSP taxa of snow and stream water communities is very apparent. In particular, the exceptionally high PSP shown by the *Methylobacterium*-1 OTU is striking for both habitat types, with other related OTUs (*Methylobacterium*-6342 and *Methylobacterium*-1508) showing very high PSP. *Methylobacterium*-1 is well represented within the snowpack 16S rRNA gene profiles of week 1, but then shows disproportionately high PSP in following weeks before its loss from the snowpack community by the fifth week. This would suggest continued protein synthesis potential is maintained for some time in spite of its rapidly diminished population size within the snowpack.

Combined, this pattern of high PSP taxa affiliated to the genus *Methylobacterium* merits further consideration. *Methylobacterium* are well known facultative methylotrophs (Chistoserdova et al 2003, Chistoserdova 2011), raising the prospect of methyl metabolism as a hitherto unappreciated metabolic strategy on the Greenland Ice Sheet. Redeker et al

(Redeker et al 2017) provided direct evidence of trace gas metabolism in polar snowpacks through the cycling of methyl halides and dimethyl sulphides. Although Redeker et al (2017) did not explore the diversity of microbes associated with methyl cycling, it is possible that *Methylobacterium* in the decaying snowpacks of the Dark Zone are involved in cycling of climate-relevant trace gases.

The relevance of disproportionately high PSP *Methylobacterium* to biogeochemical cycling in the Dark Zone is further amplified when their status as the dominant 16S rRNA (cDNA) taxon in meltwater exports is considered. Supraglacial meltwater from the Dark Zone is typically routed via surface streams into moulins to the bed of the Greenland Ice Sheet, an environment conducive for methane cycling(Yang and Smith 2016). In catchments fed by meltwater from the Dark Zone, variable rates of methanogenesis and methane oxidation have been observed, with potential impacts for the global methane cycle (Dieser et al 2014, Lamarche-Gagnon et al 2019). The discharge of highly oxygenated supraglacial meltwater into the subglacial environment is strongly associated with the cessation of methanogenesis and consumption of methane via aerobic oxidation (Dieser et al 2014). The export of high PSP *Methylobacterium* from the Dark Zone surface in this meltwater, as detected here, raises the prospect that surface-derived taxa inoculating the bed of the Greenland Ice Sheet play a role within a subglacial consortium nourished by the oxidation of methane. Further investigations focused on the fate of supraglacial microbiota transferred to the subglacial ecosystem could reveal whether this process is sufficient to mitigate the subglacial synthesis of this potent greenhouse gas.

## Conclusions

In summary, 16S rRNA gene and rRNA (cDNA) quantification and sequencing of snow, cryoconite and stream water bacterial communities from the Dark Zone of the Greenland Ice

Sheet was conducted at weekly intervals during the melt season of 2014. Recently, attention has been focused on the albedo-reducing properties of microbial consortia within the Dark Zone (Williamson et al 2019) highlighting the importance of microbial interactions in the future of the Greenland Ice Sheet. Our study addresses the related question of microbial community dynamics, and reveals that rare taxa appear to be disproportionately active. Notably, these taxa appear central to the structure of their communities and may play under-appreciated roles within the carbon cycle of the Greenland Ice Sheet. The presence of high-PSP rare taxa within *Methylobacterium* in melting snow and stream-water raises the prospect of supraglacial methyl compound cycling and export to the subglacial ecosystem. Our study represents a targeted locus amplicon sequencing approach, which in future could be complemented with genome-resolved metagenomics and direct process measurements of carbon cycling and export in both Dark Zone surface and connected downstream habitats. This would further elucidate the connections between these communities, climate change and impacts on downstream riverine and marine ecosystems from the most expansive supraglacial bare ice habitat in the Northern Hemisphere.

## ACKNOWLEDGEMENTS

This manuscript is dedicated to the memory of Kathi Hell (1985-2019) whose earlier work with us on dynamic Arctic snow microbiomes helped inspire the study described herein. Fieldwork and laboratory analyses were supported by Royal Society grant RG130314 to AE and TI-F while JKG was supported by a South African National Research Foundation Fellowship. Sequencing was performed using BBSRC funded facilities for the analyses described herein. KAC acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No

663830. Financial support was also provided to KAC by the Welsh Government and Higher Education Funding Council for Wales through the Sêr Cymru National Research Network for Low Carbon, Energy and Environment. AE is grateful for Leverhulme Trust Research Fellowship RF-2017-652\2 and NERC NE/S1001034/1 which eased completion of the work.

**The authors declare no conflict of interest**

## LIST OF FIGURES

**FIGURE 1**: Overview of the study location and physical conditions. (A) The Dark Zone of the Greenland Ice Sheet. RGB composite image of the Kangerlussuaq region of the Greenland Ice Sheet generated from the European Space Agency Sentinel 2 reflectance product (atmospheric correction and reprojection of Level 1C tile downloaded from earthexplorer.usgs.gov using Sen2Cor, then bands 2, 3 and 4 merged and scaled using GDAL) showing the study site marked with a star, 38 km inland of the Greenland Ice Sheet margin. (B) Shortwave Incident Radiation (SWIR), Long Wave Radiation (LWR), air temperature, humidity, wind speed and direction, turbulent radiation (Turb Rad) and melt rate monitored at automatic weather station site S6. Meteorological data courtesy of CJPP Smeets and MR van den Broeke, Utrecht University.

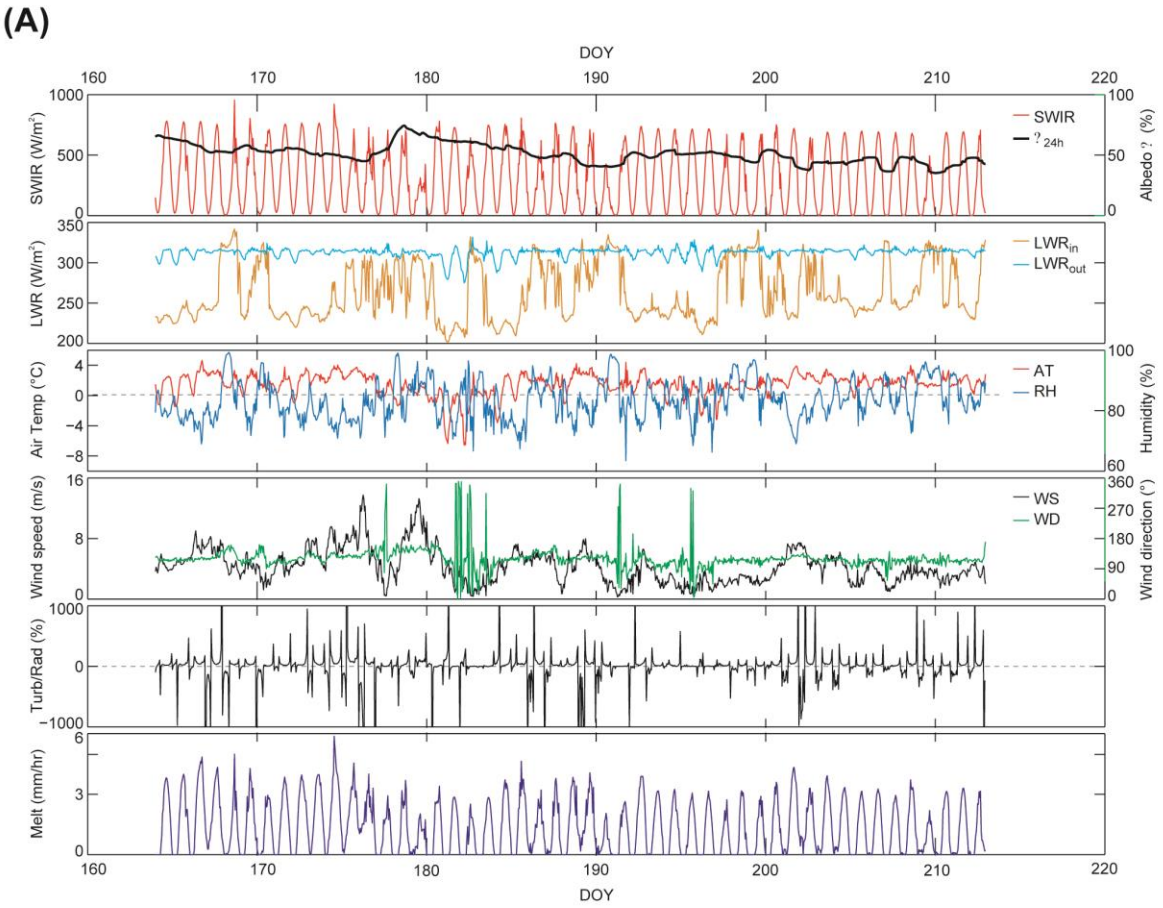
**FIGURE 2:** Quantitative PCR data on 16S rRNA gene (grey bars) and 16S rRNA (cDNA; open bars) amplifiable copy number and bacterial protein synthesis potential (diamonds) for (A) cryoconite, (B) snow, and (C) meltwater based upon analysis of triplicate weekly samples for each habitat for 7 weeks after the 19$^{th}$ of June 2014. Note the different scales for each sub-panel.

**FIGURE 3:** Overview of amplicon sequencing data. (A) Non-metric multi-dimensional scaling (nMDS) of fourth-root transformed Bray Curtis distances in OTU relative abundances

for 16S rRNA genes and 16S rRNA ordinated by habitat. (B) Community composition based on phylum (or proteobacterial class) relative abundance for 16S rRNA genes and 16S rRNA ordinated by habitat. Blank bars indicate samples excluded upon rarefaction.

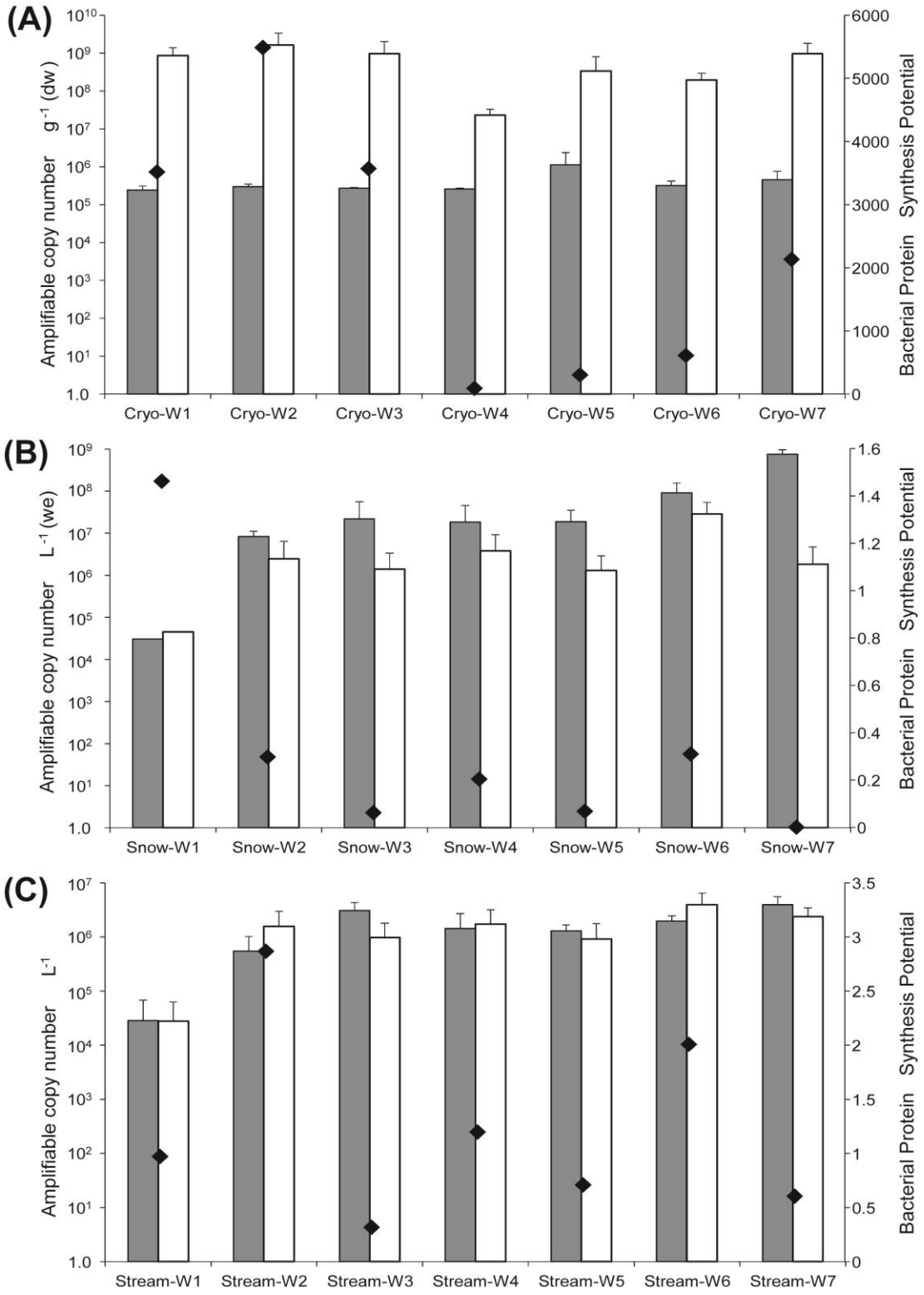
**FIGURE 4:** Relationships between 16S rRNA gene and 16S rRNA (cDNA) data for (A) cryoconite, (B) snow, and (C) streams during the study period. For each habitat, the correlation between phylum (and proteobacterial class) relative abundances between 16S rRNA gene and 16S rRNA (cDNA) data is shown. The diagonal line is used to indicate 1:1 equivalence. Rank abundance curves denote the OTU level distribution of taxa. The black line indicates the rank abundance curve of 16S rRNA gene OTUs against their relative abundance (left vertical axis). Data points represent individual OTUs, coloured by their parent phylum or proteobacterial class. Vertical lines indicate the position of taxa below 1% (red) and 0.1% (purple) relative abundance (RA).

**FIGURE 5:** Temporal dynamics of OTUs with average weekly Protein Synthesis Potential ≥1 for (A) cryoconite, (B) snow, and (C) streams during the study period. The left hand plots show the relative abundance of each taxon by week while the right hand plots show the protein synthesis potential. Each OTU is named according to the lowest grade Greengenes taxon assigned and its denovo-OTU reference.



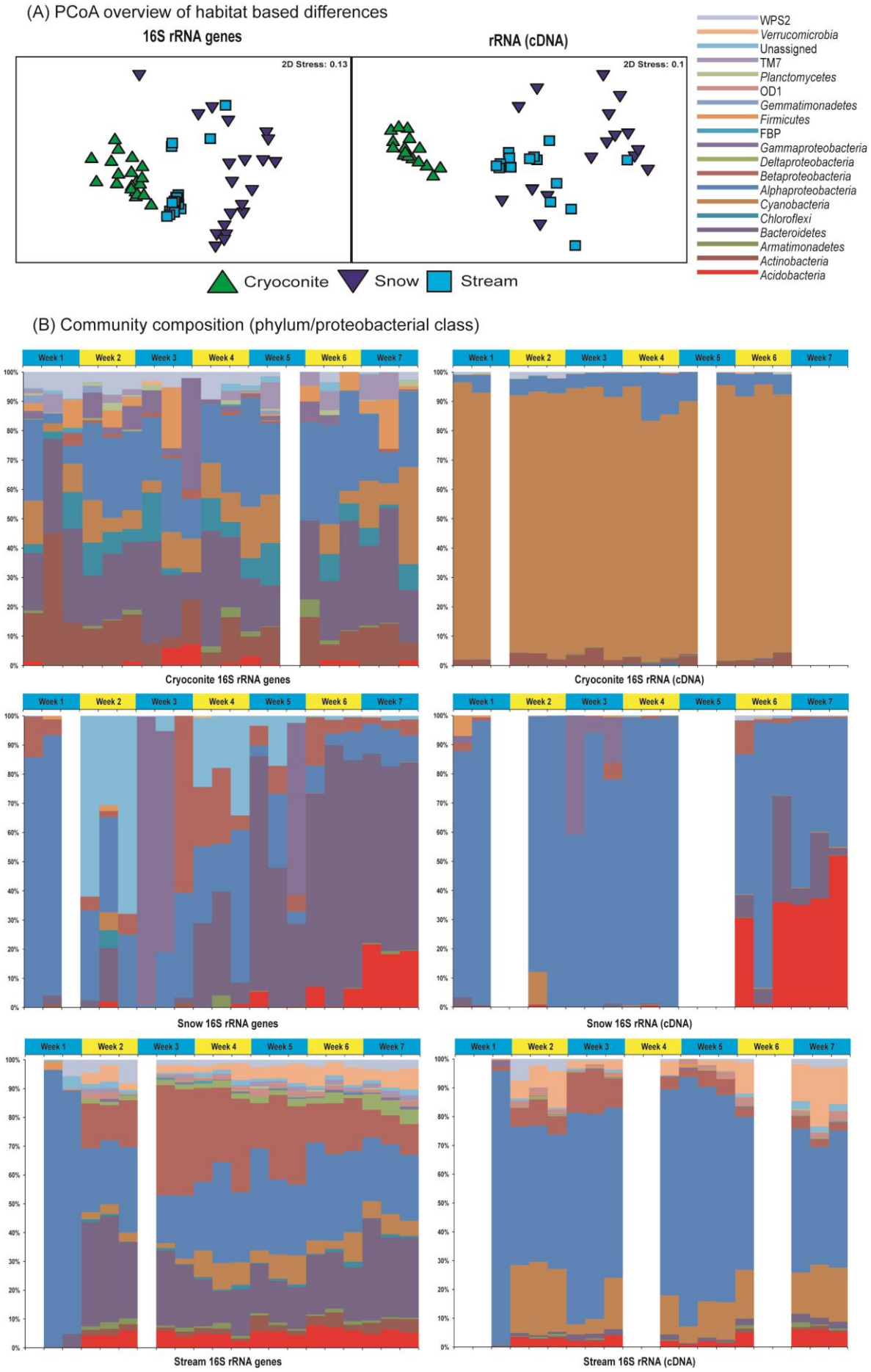

519

520

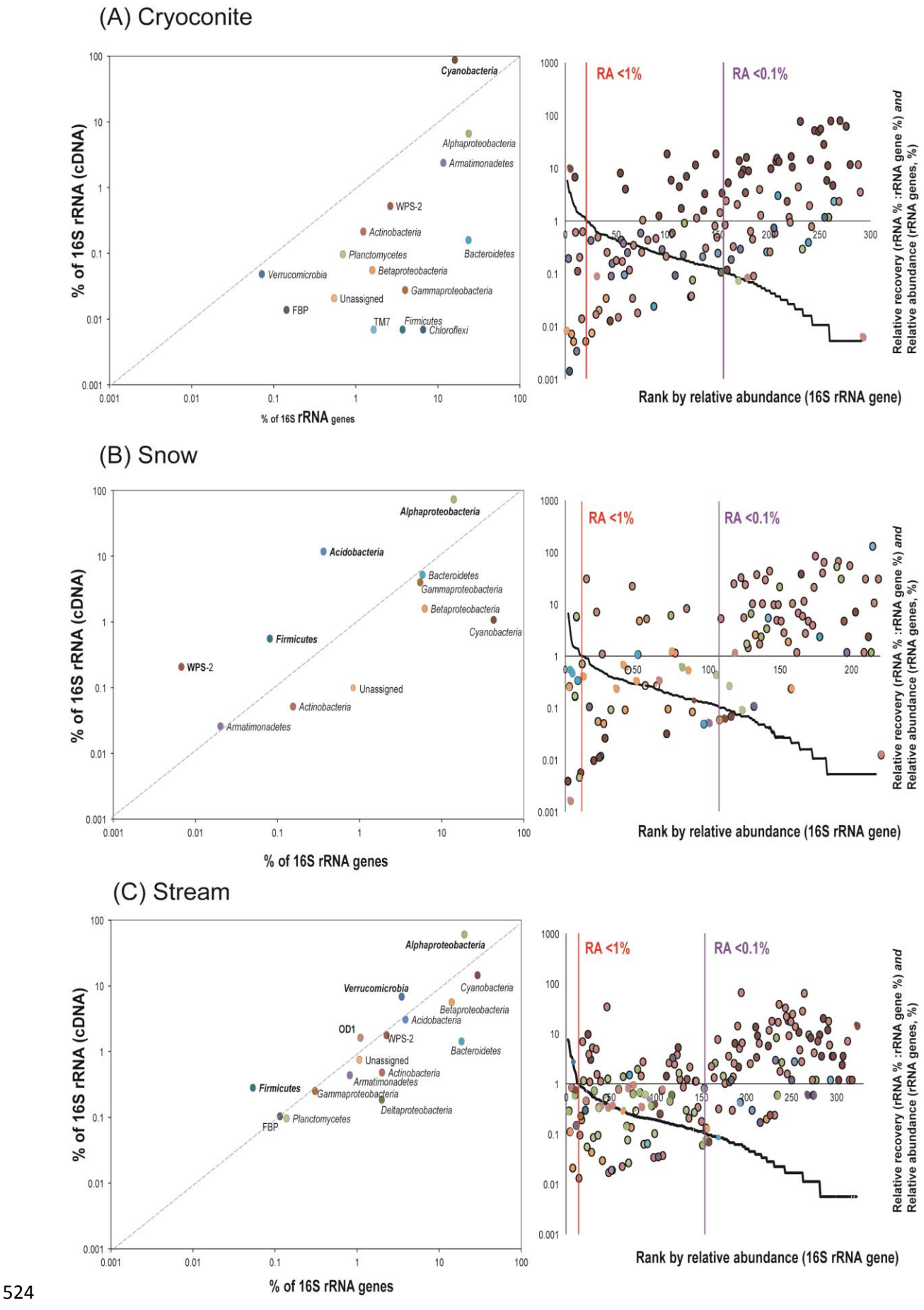



521

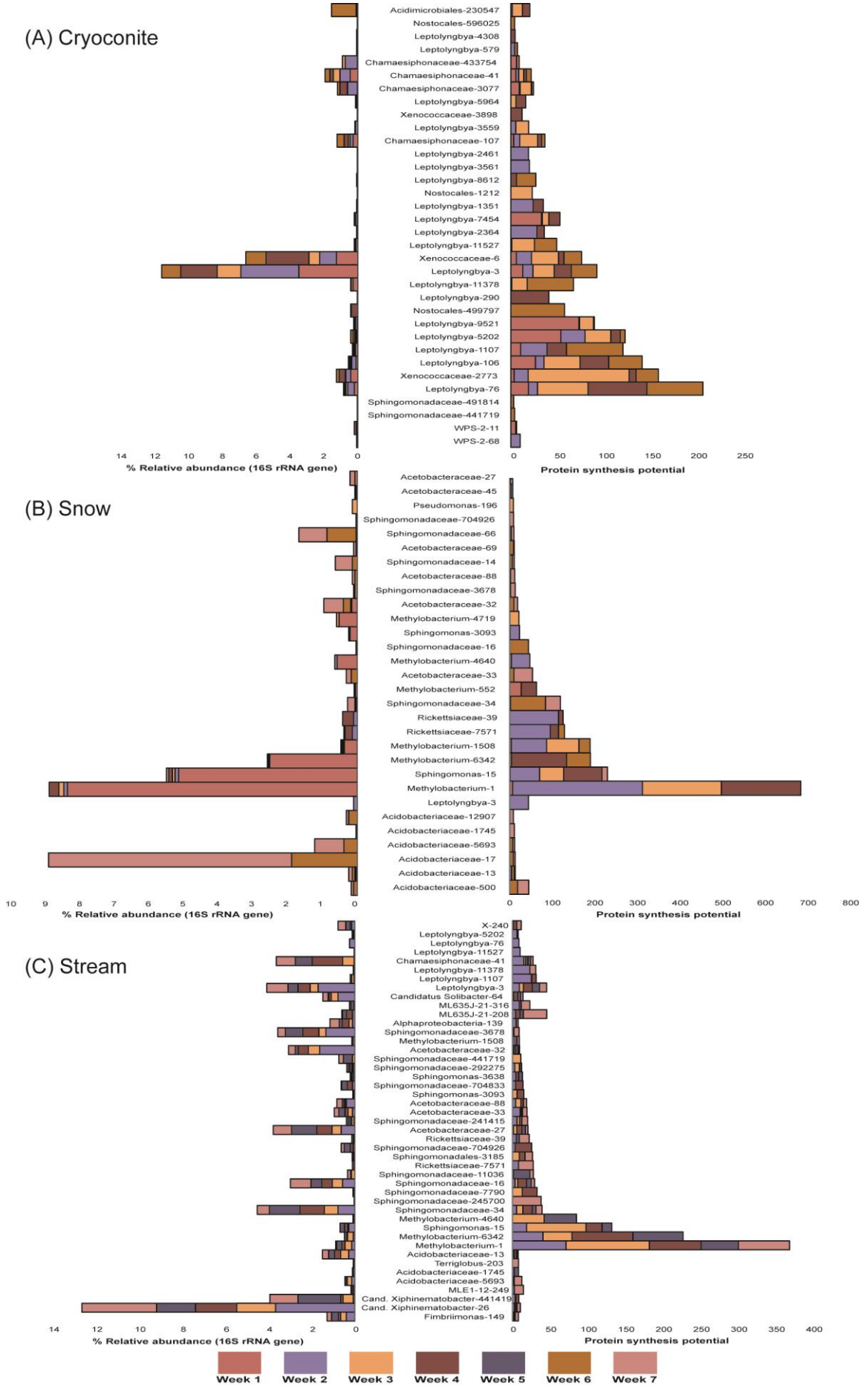
522




523




524




525

**Table 1: Cumulative Protein Synthesis Potential (PSP) and relatives of keystone OTUs present in 16S rRNA gene profiles of cryoconite (Cryo), snow and stream habitats**

| Greengenes-OTU | Phylum | Betweenness | Cryo-PSP | Snow-PSP | Stream-PSP | Relative | Taxon | Accession | % ID | Habitat |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Sphingomonadaceae*-14** | Proteobacteria | 7.17 | 1.59 | **10.54** | 3.93 | CER | Uncultured bacterium clone Bysf-47-Sf10-014 | HQ622730.1 | 99 | Austre Lovenbreen, Svalbard - glacier ice core |
| | | | | | | CNR | *Sphingopyxis* sp. JJ2203 | JX304649.1 | 98 | Jangsu-gun, South Korea - lake |
| ***Comamonadaceae*-5** | Proteobacteria | 6.25 | 0.25 | **1.89** | 1.30 | CER | *Polaromonas* sp. MDB2-14 | JX949585.1 | 99 | China - glacier |
| | | | | | | CNR | *Polaromonas* sp. MDB2-14 | JX949585.1 | 99 | China - glacier |
| ***Methylobacterium*-1** | Proteobacteria | 4.99 | - | **405.88** | 368.26 | CER | Marine bacterium MSC10 | EU753147.1 | 99 | Nahant, Massachusetts - tidal flat sand |
| | | | | | | CNR | Marine bacterium MSC10 | EU753147.1 | 99 | Nahant, Massachusetts - tidal flat sand |
| ***Sphingomonadaceae*-34** | Proteobacteria | 4.75 | 4.63 | **118.83** | 39.16 | CER | Uncultured bacterium clone QA4_1_042 | LC076727.1 | 99 | Qaanaaq Glacier, Greenland - cryoconite granules |
| | | | | | | CNR | *Novosphingobium* sp. STM-24 | LN890294.1 | 96 | Freshwater |
| ***Acetobacteraceae*-27** | Proteobacteria | 2.92 | 0.50 | 7.00 | **20.32** | CER | Uncultured bacterium clone Ms-10-Fx11-2-091 | AB990033.1 | 99 | USA:Alaska - environmental_sample |
| | | | | | | CNR | Acetobacteraceae bacterium LX45 | KC921158.1 | 99 | Ginger foundation - soil |
| ***Xenococcaceae*-6** | Cyanobacteria | 2.88 | **76.36** | 2.00 | - | CER | Uncultured cyanobacterium clone FQSS103 | EF522323.1 | 99 | Rocky Mountain - endolithic sandstone community |
| | | | | | | CNR | *Phormidium* sp. CCALA 726 | GQ504036.1 | 92 | Ny-Ålesund, Svalbard, Arctic |
| ***Xenococcaceae*-2773** | Cyanobacteria | 2.54 | **159.23** | - | 0.25 | CER | Uncultured cyanobacterium clone FQSS103 | EF522323.1 | 98 | Rocky Mountain, USA - endolithic sandstone community |
| | | | | | | CNR | *Chroococcus* sp. VP2-07 | GQ504036.1 | 91 | Italy and Spain - fountains |
| ***Sphingomonas*-15** | Proteobacteria | 2.15 | - | **147.50** | 132.00 | CER | Uncultured bacterium clone Bysf-47-Sf10-014 | KP296188.1 | 99 | Antarctic - surface seawater |
| | | | | | | CNR | *Sphingomonas* sp. UYEF32 | KU060875.1 | 99 | King George Island, Antarctica - exfoliation rock |
| ***Sphingomonadaceae*-16** | Proteobacteria | 1.83 | 3.61 | **44.00** | 29.32 | CER | Uncultured bacterium clone Bysf-47-Sf10-014 | JX967335.1 | 99 | Norway - granite outcrop |
| | | | | | | CNR | *Blastomonas* sp. TW1 | AY704922.1 | 96 | Biofilm - drinking water |
| ***Methylobacterium*-1508** | Proteobacteria | 1.38 | - | **184.50** | 9.00 | CER | Uncultured bacterium clone LIB079_B_C08 | KM852235.1 | 99 | Biofilm – drinking water |
| | | | | | | CNR | *Methylobacterium brachiatum* strain ZJ0902B96 | KU173699.1 | 98 | East China sea - surface seawaters |
| ***Leptolyngbya*-290** | Cyanobacteria | 1.09 | **41.00** | - | - | CER | Uncultured cyanobacterium clone H-D14 | DQ181732.1 | 99 | East Antarctic lakes - microbial mat |
| | | | | | | CNR | *Phormidesmis* sp. LD30 5700 TP | LN849930.1 | 98 | Western Himalaya - biological soil crust |
| ***Acetobacteraceae*-32** | Proteobacteria | 1.08 | - | **18.40** | 9.23 | CER | Uncultured bacterium QA4_30_153 | LC076735 | 98 | Cryoconite, Qaanaaq, Greenland |
| | | | | | | CNR | *Acetobacter pasteurianus* NBRC 3225 | AB680032.1 | 93 | Culture collection |
| ***Leptolyngbya*-3** | Cyanobacteria | 0.75 | **92.80** | 44.00 | 45.35 | CER | Uncultured cyanobacterium clone LH16_269 | KM112118.1 | 99 | McMurdo Dry Valley Lakes, Antarctica - benthic mats |
| | | | | | | CNR | *Phormidesmis priestleyi* ANT.L66.1 | AY493581.1 | 99 | Antarctic lake - benthic microbial mat |
| ***Leptolyngbya*-106** | Cyanobacteria | 0.51 | **141.64** | - | 6.67 | CER | Uncultured bacterium clone IT2-66 | KX247359.1 | 98 | Zhadang, China - glacier forefield |
| | | | | | | CNR | *Phormidesmis priestleyi* ANT.LG2.4 | AY493580.1 | 98 | Antarctic lake - benthic microbial mat |
| ***Leptolyngbya*-76** | Cyanobacteria | 0.51 | **207.39** | - | 8.43 | CER | Cyanobacterium cWHL-1 | HQ230236.1 | 99 | Ward Hunt Island, Ward Hunt Lake, Canada - snow |
| | | | | | | CNR | *Phormidesmis priestleyi* ANT.LG2.4 | AY493580.1 | 98 | Antarctic lake - benthic microbial mat |
| ***Methylobacterium*-6342** | Proteobacteria | 0.20 | - | 185.00 | **226.77** | CER | Uncultured alphaproteobacterium clone CN-2_B05 | EF219937.1 | 99 | Antarctica, unvegetated soil environments at Coal Nunatak |
| | | | | | | CNR | *Methylobacterium tardum* strain IHBB 11162 | KR085941.1 | 99 | Suraj Tal, Trans Himalayas - Lake Water |

REFERENCES

Anesio AM, 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:** 32-40.

Bhatia MP, Das SB, Xu L, Charette MA, Wadham JL, Kujawinski EB (2013a). Organic carbon export from the Greenland ice sheet. *Geochimica et Cosmochimica Acta* **109:** 329-344.

Bhatia MP, Kujawinski EB, Das SB, Breier CF, Henderson PB, Charette MA (2013b). Greenland meltwater as a significant and potentially bioavailable source of iron to the ocean. *Nature Geoscience* **6:** 274-278.

Björkman MP, Zarsky JD, Kühnel R, Hodson A, Sattler B, Psenner R (2014). Microbial cell retention in a melting High Arctic snowpack, Svalbard. *Arctic, Antarctic, and Alpine Research* **46:** 471-482.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME Journal* **7:** 2061-2068.

Budyko MI (1969). The effect of solar radiation variations on the climate of the earth. *Tellus* **21:** 611-619.

Cameron KA, Hagedorn B, Dieser M, Christner BC, Choquette K, Sletten R *et al* (2014). Diversity and potential sources of microbiota associated with snow on western portions of the Greenland Ice Sheet. *Environmental Microbiology***: 17**: 594-609.

Cameron KA, Stibal M, Zarsky JD, Gözdereliler E, Schostag M, Jacobsen CS (2015). Supraglacial bacterial community structures vary across the Greenland ice sheet. *FEMS microbiology ecology* **92:** fiv164.

Cameron KA, Stibal M, Hawkings JR, Mikkelsen AB, Telling J, Kohler TJ *et al* (2017). Meltwater export of prokaryotic cells from the Greenland ice sheet. *Environmental microbiology* **19:** 524-534.

Chistoserdova L, Chen S-W, Lapidus A, Lidstrom ME (2003). Methylotrophy in Methylobacterium extorquens AM1 from a genomic point of view. *Journal of Bacteriology* **185:** 2980-2987.

Chistoserdova L (2011). Modularity of methylotrophy, revisited. *Environmental Microbiology* **13:** 2603-2622.

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.

Cook JM, Hodson AJ, Anesio AM, Hanna E, Yallop M, Stibal M *et al* (2012). An improved estimate of microbially mediated carbon fluxes from the Greenland ice sheet. *Journal of Glaciology* **58:** 1098-1108.

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.

Cook JM, Edwards A, Bulling M, Mur LA, Cook S, Gokul JK *et al* (2016). Metabolome-mediated biocryomorphic evolution promotes carbon fixation in Greenlandic cryoconite holes. *Environmental Microbiology* **18:** 4674-4686.

Cook JM, Hodson AJ, Gardner AS, Flanner M, Tedstone AJ, Williamson C *et al* (2017). Quantifying bioalbedo: a new physically based model and discussion of empirical methods for characterising biological influence on ice and snow albedo. *The Cryosphere* **11:** 2611-2632.

Dieser M, Broemsen ELJE, Cameron KA, King GM, Achberger A, Choquette K *et al* (2014). Molecular and biogeochemical evidence for methane cycling beneath the western margin of the Greenland Ice Sheet. *ISME Journal*. **8:** 2305-2316**.**

Dubnick A, Kazemi S, Sharp M, Wadham J, Hawkings J, Beaton A *et al* (2017). Hydrological controls on glacially exported microbial assemblages. *Journal of Geophysical Research: Biogeosciences* **122:** 1049-1061.

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. *ISME Journal* **5:** 150-160.

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

Edwards A, Mur LAJ, Girdwood SE, Anesio AM, Stibal M, Rassner SME *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.

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:** 3752-3767.

Hawkings J, Wadham J, Tranter M, Telling J, Bagshaw E, Beaton A *et al* (2016). The Greenland Ice Sheet as a hot spot of phosphorus weathering and export in the Arctic. *Global Biogeochemical Cycles* **30:** 191-210.

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, Anesio AM, Ng F, Watson R, Quirk J, Irvine-Fynn T *et al* (2007). A glacier respires: Quantifying the distribution and respiration CO(2) flux of cryoconite across an entire Arctic supraglacial ecosystem. *Journal of Geophysical Research-Biogeosciences* **112**.

Hodson A, Boggild C, Hanna E, Huybrechts P, Langford H, Cameron K *et al* (2010). The cryoconite ecosystem on the Greenland ice sheet. *Annals of Glaciology* **51:** 123-129.

IPCC (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Change IPoC (ed). IPCC: Geneva, Switzerland.

Lamarche-Gagnon G, Wadham JL, Lollar BS, Arndt S, Fietzek P, Beaton AD *et al* (2019). Greenland melt drives continuous export of methane from the ice-sheet bed. *Nature* **565:** 73.

Langford H, Hodson A, Banwart S, Bøggild C (2010). The microstructure and biogeochemistry of Arctic cryoconite granules. *Annals of Glaciology* **51:** 87-94.

Lennon JT, Jones SE (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology* **9:** 119-130.

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

Lutz S, McCutcheon J, McQuaid JB, Benning LG (2018). The diversity of ice algal communities on the Greenland Ice Sheet as revealed by oligotyping. *Microbial Genomics* **4:** - e000159.

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

Musilova M, Tranter M, Wadham J, Telling J, Tedstone A, Anesio AM (2017). Microbially driven export of labile organic carbon from the Greenland ice sheet. *Nature Geoscience* **10:** 360.

Nordenskiöld AE (1870). *Redogörelse för en expedition til Grönland År 1870*. P.A. Norstedt & Söner: Stockholm.

Pedrós-Alió C (2012). The rare bacterial biosphere. *Annual Review of Marine Science* **4:** 449-466.

Peura S, Bertilsson S, Jones RI, Eiler A (2015). Resistant Microbial Cooccurrence Patterns Inferred by Network Topology. *Applied and Environmental Microbiology* **81:** 2090-2097.

Power ME, Mills LS (1995). The keystone cops meet in Hilo. *Trends in Ecology & Evolution* **10:** 182-184.

Redeker KR, Chong JPJ, Aguion A, Hodson A, Pearce D (2017). Microbial metabolism directly affects trace gases in (sub) polar snowpacks. *Journal of The Royal Society Interface* **14:** 20170729.

Ryan J, Smith L, van As D, Cooley S, Cooper M, Pitcher L *et al* (2019). Greenland Ice Sheet surface melt amplified by snowline migration and bare ice exposure. *Science Advances* **5:** eaav3738.

Ryan JC, Hubbard A, Stibal M, Box JE, team TDSP (2016). Attribution of Greenland's ablating ice surfaces on ice sheet albedo using unmanned aerial systems. *The Cryosphere Discuss* **2016:** 1-23.

Ryan JC, Hubbard A, Stibal M, Irvine-Fynn TD, Cook J, Smith LC *et al* (2018). Dark zone of the Greenland Ice Sheet controlled by distributed biologically-active impurities. *Nature communications* **9:** 1065.

Segawa T, Yonezawa T, Edwards A, Akiyoshi A, Tanaka S, Uetake J *et al* (2017). Biogeography of cryoconite forming cyanobacteria on polar and Asian glaciers. *Journal of biogeography* **44:** 2849-2861.

Smith HJ, Schmit A, Foster R, Littman S, Kuypers MM, Foreman CM (2016). Biofilms on glacial surfaces: hotspots for biological activity. *NPJ Biofilms and Microbiomes* **2:** 16008.

Smith LC, Yang K, Pitcher LH, Overstreet BT, Chu VW, Rennermalm ÅK *et al* (2017). Direct measurements of meltwater runoff on the Greenland ice sheet surface. *Proceedings of the National Academy of Sciences***:** 201707743.

Steven B, Hesse C, Soghigian J, Dunbar J (2017). Simulated Ribosomal RNA: DNA Ratios Show Potential to Misclassify Active Populations as Dormant. *Applied and environmental microbiology***:** AEM. 00696-00617.

Stibal M, Lawson EC, Lis GP, Mak KM, Wadham JL, Anesio AM (2010). Organic matter content and quality in supraglacial debris across the ablation zone of the Greenland ice sheet. *Annals of Glaciology* **51:** 1-8.

Stibal M, Schostag M, Cameron KA, Hansen LH, Chandler DM, Wadham JL *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:** 293-300.

Stibal M, Box JE, Cameron KA, Langen PL, Yallop ML, Mottram RH *et al* (2017). Algae drive enhanced darkening of bare ice on the Greenland ice sheet. *Geophysical Research Letters* **44**. 11,463-11,471.

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.

Tedstone AJ, Bamber JL, Cook JM, Williamson CJ, Fettweis X, Hodson AJ *et al* (2017). Dark ice dynamics of the south-west Greenland Ice Sheet. *Cryosphere* **11:** 2491-2506.

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:** fiw127.

Van Tricht K, Lhermitte S, Lenaerts JT, Gorodetskaya IV, L'Ecuyer TS, Noël B *et al* (2016). Clouds enhance Greenland ice sheet meltwater runoff. *Nature Communications* **7:** 10266.

Wientjes IGM, Oerlemans J (2010). An explanation for the dark region in the western melt zone of the Greenland ice sheet. *Cryosphere* **4:** 261-268.

Wientjes IGM, Van de Wal RSW, Reichart GJ, Sluijs A, Oerlemans J (2011). Dust from the dark region in the western ablation zone of the Greenland ice sheet. *Cryosphere* **5:** 589-601.

Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA, Quince C *et al* (2014). Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environ Microbiol* **16:** 2514-2524.

Williamson C, Anesio A, Cook J, Tedstone A, Poniecka E, Holland A *et al* (2018). Ice algal bloom development on the surface of the Greenland Ice Sheet. *FEMS microbiology ecology* **94:** fiy025.

Williamson C, Cameron KA, Cook JM, Zarsky JD, Stibal M, Edwards A (2019). Glacier algae: a dark past and a darker future. *Frontiers in microbiology* **10:** 524.

Yallop ML, Anesio AM, Perkins RG, Cook J, Telling J, Fagan D *et al* (2012). Photophysiology and albedo-changing potential of the ice algal community on the surface of the Greenland ice sheet. *ISME J* **6:** 2302-2313.

Yang K, Smith LC (2016). Internally drained catchments dominate supraglacial hydrology of the southwest Greenland Ice Sheet. *Journal of Geophysical Research: Earth Surface* **121:** 1891-1910.

# Supplementary Methods for: Illuminating the functional rare biosphere of the Greenland Ice Sheet's Dark Zone

## Sample archiving

Other researchers have documented the half-life of the cellular RNA pool (which is heavily dominated by rRNA e.g. Moran *et al.* 2012) of glacial bacteria at the low temperatures typical of ice surfaces is less than one day (Segawa *et al.* 2014), with the implication that rRNA turnover rates are sufficient at the temporal resolution of this study to capture meaningful changes in protein synthesis potential well within the expected community doubling time (Anesio *et al.* 2010). However, to prevent sample degradation, this necessitates careful sample archiving to stabilize the rRNA and genomic DNA pools of the collected biomass. Therefore, due to the remote location of the field camp, in line with established procedures (e.g. Stibal *et al.* 2015, Cameron *et al.* 2015), Soil Lifeguard RNA/DNA preservative (MO BIO Laboratories) stabilized samples were stored dark, within crushed ice, prior to transfer to -20°C storage at the Kangerlussuaq International Science Support field laboratory, within three weeks of collection, followed by -80°C archiving at the Aberystwyth laboratory until further processing within four months.

Specifically, in the field, cryoconite samples were aspirated into sterile microcentrifuge tubes and immersed immediately in Soil Lifeguard RNA/DNA preservative (MO BIO Laboratories). Snow samples representing 1.2-1.5L water equivalent were collected in sterile whirlpaks (Nasco, Inc.) and melted in the dark at +10°C while meltwater from supraglacial stream meltwater (ca. 2L) was collected in whirlpaks. Melted snow was filtered through 0.22 µm Sterivex GP polyethersulfone filters (Millipore) and chemically preserved within 12 hours of sampling, and stream water was filtered and chemically preserved within 6 hours of sampling. The Sterivex filters were connected to tubing which was thoroughly rinsed using 10% HCl and 70% ethanol to remove potential contaminants from the tubing entering the whirpak bag. Sterivex filters were filled with 2 mL Soil Lifeguard RNA/DNA preservative (MO BIO Laboratories).

Nucleic Acid Extraction

All pre-PCR procedures were conducted aseptically in a bleach-disinfected laminar flow hood with certified DNA/RNA free plasticware including aerosol resistant tips. Negative extraction controls were prepared using sterivex cartridges and blank cryoconite extraction.

Total nucleic acids were co-extracted from cryoconite using the PowerBiofilm™ RNA Isolation Kit (MO BIO Laboratories) according to the manufacturer's instructions. 0.2g ±0.05g of wet sediment was used in each extraction. Nucleic acids from snow and stream samples were co-extracted using the PowerWater® Sterivex™ RNA Kit (MO BIO Laboratories) according to the manufacturer's protocol with minor adjustments for DNA extraction recommended by MO BIO Laboratories (Dr Emelia DeForce, personal communication). Specifically, these steps required the addition of 20 µL beta-mercaptoethanol for every 880 µL of solution ST1B, the additional incubation of the sterivex

filter at 70°C for 10 minutes during lysis and the addition of an equal volume of 100% ethanol to buffer ST4. All RNA samples were subjected to DNase treatment using the RTS DNase™ Kit (MO BIO Laboratories) according to the manufacturer's instructions. Aliquots of each RNA sample were then used as template in first strand cDNA synthesis via reverse transcription with SuperScript™ III Reverse Transcriptase (Invitrogen) and universal 16S rRNA gene region primer 1389R, according to the manufacturer's instructions, for use in downstream PCR applications.

16S ribosomal RNA (cDNA) and 16S ribosomal RNA gene quantitative PCR

To estimate 16S rRNA gene and 16S rRNA abundance within the samples, two microliters of each DNA extract or reverse transcriptase product was used as template in 20 µL reactions consisting of 1× SensiFast SYBR (no ROX) qPCR mixture (BioLine, Ltd) and 16S rRNA primers 27F (5'-AGAGTTTGATCCTGGCTCAG) and 1389R (5'-ACGGGCGGTGTGTACAAG) amplified for 35 cycles in a Mic real time PCR cycler (BioMolecularSystems, Inc.) The copy number of gene and 16S rRNA molecules was estimated from a plasmid-borne 27F-1389R standard, and converted to gram dry weight cryoconite or meltwater stream volume or water equivalent volume of snow (by reference to the volume of samples filtered) respectively. We note the uncertainties associated with extraction efficiencies for different microbial taxa within environmental samples subjected to nucleic acid extraction, therefore we present the copy number of 16S rRNA genes and 16S rRNA as the *amplifiable* copy number per unit sample.

16S ribosomal RNA gene amplicon sequencing

Cryoconite, snow and stream DNA and RNA samples were prepared for paired end MiSeq sequencing (Illumina). Using a 3 stage PCR method, the bacterial 16S rRNA V3-V4

hypervariable region was amplified for the attachment of Nextera XT dual indices and Illumina sequencing adapters.

The primers used for 16S rRNA gene specific amplification were universal 16S primers 27F and 1389R using cycling conditions of 95 °C for 5 mins (initial denaturation); 30 cycles: 95 °C for 30 secs (denaturation), 60 °C for 30 secs (annealing), 72 °C for 45 secs (extension); 72 °C for 7 mins (final extension). Amplicons produced were used as template for nested PCRs with Nextera primers (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' and 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3') (Klindworth et al, 2013) in 12.5µL Platinum Taq Mix PCR reactions, as per manufacturer's instructions, using cycling conditions of 95°C for 3 mins (initial denaturation); 25 cycles: 95°C for 30s (denaturation), 55°C for 30s (annealing), 72°C for 30s (extension); 72°C for 5 mins (final extension). Amplification was confirmed by gel electrophoresis on 1% agarose before products were used as template in the third PCR using Platinum Taq Mix and the Nextera XT Index Kit (Illumina) to obtain unique combinations of the N7 and S5 primer barcodes per sample (SUPPLEMENTARY TABLE 1). Cycling conditions were 95° for 3 mins (initial denaturation); 8 cycles of 95°C for 30s (denaturation), 55°C for 30s (annealing), 72°C for 30s (extension); 72°C for 5 mins (final extension). Amplicons were purified using the Montage™ PCR plate filtration system (Millipore) prior to quantification on an Epoch™ spectrophotometer (BioTek). Products were then normalised to the lowest common concentration for each library (DNA 5ng/µL; cDNA 0.8ng/µL) and pooled before electrophoresis on 1% agarose gel for purification via gel extraction using the Qiagen Gel Extraction Kit (Qiagen) as per manufacturer's instructions. Libraries were sequenced at the IBERS Translational Sequencing Facility (IBERS Phenomics Centre, Gogerddan, Aberystwyth, Ceredigion SY23 3EE) on the Illumina MiSeq® platform (Illumina Inc., San Diego, CA, USA) using the MiSeq® Reagent Kit, version 3 (Illumina Inc., San Diego, CA,

USA). Initial processing of Illumina MiSeq® sequence data was performed in BaseSpace® (Illumina Inc., San Diego, CA, USA). Data were demultiplexed and the indices removed. Reads were trimmed to remove read-through into the adaptor sequence at the 3' end. Sequence files were generated in the fastq format and imported into QIIME 1.9.0 for merging of paired end reads.

Sequence Processing and Analysis

Resulting sequences were quality filtered and processed in QIIME 1.9.0 (Caporaso *et al.*, 2010) using default quality filters unless otherwise stated. Paired end sequences were joined and probed for chimeras using USearch6.1 (Edgar, 2010) which were then removed. Operational taxonomic units (OTUs) were then assigned at 97% identity using a reference based UCLUST algorithm (Edgar, 2010) and the Greengenes 13_8 database, August 2013 version (DeSantis *et al.*, 2006) before OTU tables were generated, and processed for rarefaction curves and alpha diversity indices. Multivariate analysis was performed in PRIMER 6/PERMANOVA+ (PRIMER-E Ltd, Plymouth UK). Permutational multivariate analysis of variation (PERMANOVA), principal coordinates analysis (PCO) were performed as described previously (Edwards *et al.*, 2014) with Bray Curtis transformations at the fourth root used for OTU relative abundance. One-way ANOVA was performed in Minitab 17. For samples with both 16S rRNA gene and 16S rRNA profiles, protein synthesis potential (sensu Denef *et al.*, 2016) was defined as the ratio of the relative abundances of each OTU. Community analysis using pairwise Spearman correlations in R was able to identify OTUs with high *betweenness* centrality (shortest number of paths between any two other OTUs passing through that OTU) (Peura et al, 2015), high *degree* centrality (the number of associations an OTU has) and high *closeness* centrality (the average distance between two OTUs) measures (Berry and Widder, 2014). These taxa play a central role in the ecosystem and are recognised as “bottleneck or keystone taxa” (Paine, 1966, Mills et al, 1993, Cottee-

Jone and Whittaker, 2012). Sequences are available on EBI-SRA under the study accession number PRJNA318626.

**References**

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

Berry, D. and Widder, S. (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol. 5: 1–14.

Cameron, K.A., Stibal, M., Zarsky, J.D., Gözdereliler, E., Schostag, M. and Jacobsen, C.S., 2015. Supraglacial bacterial community structures vary across the Greenland ice sheet. FEMS microbiology ecology, 92(2), fiv164.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Gonzalez Pena, A., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., Mcdonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7, 335 - 336.

Cottee-Jones, Henry Eden WWhittaker, Robert J. 2012. The keystone species concept: a critical appraisal. Frontiers of Biogeography, 4(3).

Desantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G. L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol, 72, 5069-72.

Denef, V. J., Fujimoto, M., Berry, M. A. & Schmidt, M. L. 2016. Seasonal succession leads to habitat-dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. Frontiers in Microbiology.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26, 2460-2461.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic acids research, 41(1),e1-e1.

Mills, L.S., Soulé, M.E., Doak, D.F. 1993. The Keystone-Species Concept in Ecology and Conservation. BioScience, 43(4), 219–224.

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A. and Smith, C.B., 2013. Sizing up metatranscriptomics. The ISME journal, 7(2), 237.

Paine, R. T. 1966. Food Web Complexity and Species Diversity. The American Naturalist 100(910), 65-75.

Peura, S., Bertilsson, S., Jones, R. I. & Eiler, A. 2015. Resistant microbial cooccurrence patterns inferred by network topology. Applied and Environmental Microbiology, 81, 2090-7.

Segawa, T., Ishii, S., Ohte, N., Akiyoshi, A., Yamada, A., Maruyama, F., Li, Z., Hongoh, Y. and Takeuchi, N., 2014. The nitrogen cycle in cryoconites: naturally occurring nitrification-denitrification granules on a glacier. Environmental microbiology, 16(10),3250-3262.

Stibal, M., Schostag, M., Cameron, K.A., Hansen, L.H., Chandler, D.M., Wadham, J.L. and Jacobsen, C.S., 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.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol, 73, 5261-7.

# Supplementary Methods for: Illuminating the functional rare biosphere of the Greenland Ice Sheet's Dark Zone

SUPPLEMENTARY TABLE 1: Dual Index Primers for the 96 sample Nextera XT Index Kit

| **Index 1 – I7** | **SEQUENCE** |
|---|---|
| N701 | TAAGGCGA |
| N702 | CGTACTAG |
| N703 | AGGCAGAA |
| N704 | TCCTGAGC |
| N705 | GGACTCCT |
| N706 | TAGGCATG |
| N707 | CTCTCTAC |
| N708 | CAGAGAGG |
| N709 | GCTACGCT |
| N710 | CGAGGCTG |
| N711 | AAGAGGCA |
| N712 | GTAGAGGA |

| **Index 2 – I5** | **SEQUENCE** |
|---|---|
| S501 | TAGATCGC |
| S502 | CTCTCTAT |
| S503 | TATCCTCT |
| S504 | AGAGTAGA |
| S505 | GTAAGGAG |
| S506 | ACTGCATA |
| S507 | AAGGAGTA |
| S508 | CTAAGCCT |

SUPPLEMENTARY TABLE 2: Permutational analysis of variance (PERMANOVA) across 16S rRNA gene and 16S rRNA ("nucleic acid"), day of year ad cryoconite, snow,and stream habitats.

| Test | Pseudo *F* | *p* |
|---|---|---|
| Habitats | 13.796 | 0.0001 |
| Cryoconite DNA vs RNA | 22.536 | 0.0001 |
| Cryoconite DNA over time | 1.0092 | 0.4477 |
| Cryoconite RNA over time | 1.6963 | 0.0254 |
| Snow DNA vs RNA | 5.3008 | 0.0001 |
| Snow DNA over time | 2.6525 | 0.0001 |
| Snow RNA over time | 2.9937 | 0.0003 |
| Stream DNA vs RNA | 7.1009 | 0.0001 |
| Stream DNA over time | 2.4599 | 0.0041 |
| Stream RNA over time | 1.4237 | 0.1069 |

925

SUPPLEMENTARY TABLE 3: Identity and abundance of contaminating sequences detected in DNA extraction controls

| **OTU** | **Greengenes ID** | **No of reads in NTC** | **BLAST ID** | **Accession No** | **%ID** |
|---|---|---|---|---|---|
| denovo-87257 | g_*Salinibacterium* | 16 | *Leifsonia* sp. URHE0078 | LN876553.1 | 99 |
| denovo-114112 | f_*Microbacteriaceae* | 12 | *Salinibacterium amurskyense* IC A9 | KU925169.1 | 98 |
| denovo-125684 | o_*Streptophyta* | 11 | *Roya obtusa* | KU646496.1 | 97 |
| denovo-90892 | g_*Leptolyngbya* | 7 | *Phormidesmis priestleyi* ANT.L66.1 | AY493581.1 | 99 |
| denovo-96273 | g_*Kocuria* | 4 | *Kocuria subflava* | NR_144586.1 | 99 |
| denovo-67624 | f_*Sphingomonadaceae* | 4 | *Polymorphobacter* sp. Ap23E | KX990242.1 | 99 |
| denovo-106624 | f_*Microbacteriaceae* | 3 | *Salinibacterium amurskyense* IC A9 | KU925169.1 | 98 |

SUPPLEMENTARY TABLE 4: The identity of top OTUs present in cryoconite[a], snow[b] and stream[c] water contributing to 99% of relative abundance per habitat.

| CORE OTU | Closest Environmental Relative | | | | Closest Named Relative | | | |
|---|---|---|---|---|---|---|---|---|
| | CER | Accession number | % ID | Habitat | CNR | Accession number | % ID | Habitat |
| ***Streptophyta-0***[b,c] | Uncultured bacterium clone IC4001 | HQ622719.1 | 99 | Austre Lovenbreen, Svalbard - glacier surface ice | *Roya obtusa* culture-collection SAG:168.80 chloroplast, | KU646496.1 | 97 | Freshwater algae |
| ***Methylobacterium-1***[b,c] | Marine bacterium MSC10 | EU753147.1 | 99 | Nahant, Massachusetts - tidal flat sand | Marine bacterium MSC10 | EU753147.1 | 99 | Nahant, Massachusetts - tidal flat sand |
| ***Leptolyngbya-3***[a,c] | Uncultured *cyanobacterium* clone LH16_269 | KM112118.1 | 99 | McMurdo Dry Valley Lakes, Antarctica - benthic mats | *Phormidesmis priestleyi* ANT.L66.1 | AY493581.1 | 99 | Antarctic lake - benthic microbial mat |
| ***Sphingobacteriaceae-4***[a,c] | Uncultured *Actinobacterium* clone IC4008 | HQ622724.1 | 99 | Austre Lovenbreen, Svalbard - glacier ice core | *Solitalea koreensis* strain R2A36-4 | NR_044568.1 | 90 | Soil bacterium |
| ***Comamonadaceae-5***[b,c] | *Polaromonas* sp. MDB2-14 | JX949585.1 | 99 | China - glacier | *Polaromonas* sp. MDB2-14 | JX949585.1 | 99 | China - glacier |
| ***Xenococcaceae-6***[a,c] | Uncultured *cyanobacterium* clone FQSS103 | EF522323.1 | 99 | Rocky Mountain - endolithic sandstone community | *Phormidium* sp. CCALA 726 | GQ504036.1 | 92 | Ny-Ålesund, Svalbard, Arctic |
| ***Hymenobacter-7***[b] | Uncultured bacterium clone Bysf-33-Sf11-054 | AB991133.1 | 99 | Byron Glacier, USA: Alaska - glacier surface | *Hymenobacter* sp. Ht11 | JX949241.1 | 97 | China - glacier |
| ***Saprospiraceae-8***[a,c] | Uncultured bacterium clone QA48 | LC030301.1 | 100 | Qaanaaq Glacier, North Western Greenland - glacier surface ice | Filamentous bacterium Plant1 Iso8 | DQ232754.1 | 86 | Activated sludge |
| ***Sediminibacterium-9***[a] | Uncultured *Sediminibacterium* sp. clone Limnopolar-0.5-B2 | FR848659.1 | 99 | Byers Peninsula, Antarctica - limnopolar lake | *Vibrionimonas magnilacihabitans* strain MU-2 | NR_133888.1 | 99 | Lake Michigan - water sample |
| ***Methylobacterium-10***[b,c] | Uncultured bacterium clone Bysf-47-Sf10-014 | AB991014.1 | 99 | Byron Glacier, USA:Alaska - glacier surface | *Methylobacterium* sp. 63 | JF905617.1 | 99 | Barrientos Island, Antarctica - soil |
| ***Acidimicrobiales-12***[a] | Uncultured bacterium clone AM 5.5m_4_43 | HE616493.1 | 99 | Alinen Mustajärvi, Finland - Boreal Lake | *Actinobacterium* TC4 | JF510471.1 | 94 | Chalcopyrite bio-heap - leachate |
| ***Sphingomonadaceae-14***[a,c] | Uncultured bacterium clone Bysf-47-Sf10-014 | HQ622730.1 | 99 | Austre Lovenbreen, Svalbard - glacier ice core | *Sphingopyxis* sp. JJ2203 | JX304649.1 | 98 | Jangsu-gun, South Korea - lake |
| ***Sphingomonas-15***[b,c] | Uncultured bacterium clone Bysf-47-Sf10-014 | KP296188.1 | 99 | Antarctic - surface seawater | *Sphingomonas* sp. UYEF32 | KU060875.1 | 99 | King George Island, Antarctica - exfoliation rock |
| ***Sphingomonadaceae-16***[a,c] | Uncultured bacterium clone Bysf-47-Sf10-014 | JX967335.1 | 99 | Norway - granite outcrop | *Blastomonas* sp. TW1 | AY704922.1 | 96 | Biofilm - drinking water |
| ***Acidobacteriaceae-17***[b,c] | Uncultured *Acidobacteria* bacterium clone IC3076 | HQ595216.1 | 99 | Austre Lovenbreen, Svalbard - glacier surface ice | *Granulicella aggregans* strain TPB6028 | NR_115070.1 | 99 | West Russia Tomsk - sphagnum peat |
| ***Streptophyta-19***[a] | *Carum carvi* chloroplast | KR048286.1 | 99 | Unpublished | *Carum carvi* | KR048286.1 | 99 | Unpublished |

| CORE OTU | Closest Environmental Relative | | | | Closest Named Relative | | | |
|---|---|---|---|---|---|---|---|---|
| | CER | Accession number | % ID | Habitat | CNR | Accession number | % ID | Habitat |
| | | | | | chloroplast, | | | |
| ***Thermogemmatisporaceae-21***[a] | Uncultured *Ktedobacteria* bacterium clone UMAB-cl-174 | FR749799.1 | 99 | Antarctic Peninsula - soil | *Ktedonobacter racemifer* strain DSM 44963 | NR_112949.1 | 89 | Compost |
| ***ACK-M1-22***[c] | Uncultured bacterium clone QA4_1_060 | LC076728.1 | 99 | Qaanaaq Glacier, Greenland - cryoconite granules | Candidatus *Planktophila limnetica* strain MWH-EgelM2-3.acI | FJ428831.1 | 97 | Lake Grossegelsee, Austria - freshwater lake |
| ***Phyllobacterium-24***[c] | Uncultured bacterium clone PL3-9 16S | EU527093.1 | 99 | Palong No.4 glacier, Tibet China - snow | *Phyllobacterium* sp. MD10 | KF358323.1 | 99 | Tibetan Plateau - glacier snow |
| ***Candidatus Xiphinematobacter-26***[c] | Uncultured *Xiphinematobacteriaceae* bacterium clone AL5.23 | GU047442.1 | 99 | High-Arctic acidic wetland active layer soil | *Spartobacteria* bacterium Gsoil 144 | AB245342.1 | 92 | Ginseng field - soil |
| ***Acetobacteraceae-27***[c] | Uncultured bacterium clone Ms-10-Fx11-2-091 | AB990033.1 | 99 | USA:Alaska - environmental_sample | *Acetobacteraceae* bacterium LX45 | KC921158.1 | 99 | Ginger foundation - soil |
| ***Pseudomonas-28***[b] | *Pseudomonas salomonii* strain HS9-MRL | KX128926.1 | 99 | Pakistan: Siachen Glacier, ice, water and sediment | *Pseudomonas* sp. 36 | KX354891.1 | 99 | Antarctic - ornithogenic soil |
| ***Stramenopiles-29***[c] | Uncultured bacterium clone BJOH-71 | KP724732.1 | 99 | South Shetland Archipilego, Antarctica - glacier ice | *Ochromonas* sp. CCMP1393 plastid, | KJ877675.1 | 96 | Algal origin |
| ***Sphingomonas-31***[b] | Uncultured bacterium clone LIB062_A_D01 | KM851700.1 | 100 | Drinking water biofilm | *Sphingomonas* sp. UV9 | KR922276.1 | 99 | Antarctic, unpublished |
| ***Sphingomonadaceae-34***[a,b,c] | Uncultured bacterium clone QA4_1_042 | LC076727.1 | 99 | Qaanaaq Glacier, Greenland - cryoconite granules | *Novosphingobium* sp. STM-24 | LN890294.1 | 96 | Freshwater |
| ***Rickettsiaceae-39***[b] | Uncultured bacterium clone Malla4.68 | AM945518.1 | 95 | Arctic tundra - soil | *Candidatus Trichorickettsia mobilis* clone 36 | AM945518.1 | 95 | Wastewater plant |
| ***Chamaesiphonaceae-41***[a,c] | Uncultured bacterium clone EpiUMB29 | FJ849281.1 | 99 | Noatak National Preserve, Alaska - Arctic stream epilithon | *Chamaesiphon subglobosus* PCC 7430 | AY170472.1 | 98 | (culture collection) |
| ***Comamonadaceae-44***[b] | *Variovorax paradoxus* strain 11-4(2) | KT369961.1 | 99 | Qilian mountain China | *Variovorax paradoxus* strain 11-4(2) | KT369961.1 | 99 | Qilian mountain, China |
| ***Pseudomonas-46***[b] | Uncultured bacterium clone MTWL201306-93 | KX509317.1 | 99 | Rainwater | *Pseudomonas* sp. DMSP-3 | KU296913.1 | 99 | Kongsfjorden, Svalbard - Arctic seawater |
| ***Ralstonia-52***[b] | Uncultured *Ralstonia* sp. clone BF64A_B11 | HM141108.1 | 100 | Borup Fiord, Ellesmere Island, Nunavut, Canada - supraglacial spring outflow | *Ralstonia pickettii* strain HT16-MRL | KP318066.1 | 99 | Hindu Kush mountain range, Chitral - Tirich Mir glacier |
| ***Micrococcus-70***[a] | *Micrococcus* sp. strain BAB-5964 | KX622627.1 | 100 | Soil | *Micrococcus* sp. H-CD9b | KT799848.1 | 100 | Northern Patagonia, Chile - Comau Fjord |
| ***Leptolyngbya-76***[a] | *Cyanobacterium* cWHL-1 | HQ230236.1 | 97 | Ward Hunt Island, Ward Hunt Lake, Canada - snow | *Cyanobacterium* cWHL-1 | HQ230236 | 99 | Canadian High Arctic |

| CORE OTU | Closest Environmental Relative | | | | Closest Named Relative | | | |
|---|---|---|---|---|---|---|---|---|
| | CER | Accession number | % ID | Habitat | CNR | Accession number | % ID | Habitat |
| ***Leptolyngbya-106***[a] | Uncultured bacterium clone IT2-66 | KX247359.1 | 98 | Zhadang, China - glacier forefield | *Cyanobacterium* cWHL-1 | HQ230236.1 | 98 | Canadian High Arctic |
| ***Chamaesiphonaceae-107***[a] | Uncultured *cyanobacterium* clone p660_S12 | JQ407506.1 | 99 | Greenland - glacier ice | *Chroococcales cyanobacterium* CYN67 | JQ687333.1 | 98 | Antarctica/New Zealand (not published) |
| ***Leptolyngbya-290***[a] | Uncultured *cyanobacterium* clone H-D14 | DQ181732.1 | 99 | East Antarctic lakes - microbial mat | *Phormidesmis* sp. LD30 5700 TP | LN849930.1 | 98 | Western Himalaya - biological soil crust |
| ***Leptolyngbya-1107***[a] | Uncultured *cyanobacterium* clone LH16_269 | KM112118.1 | 99 | McMurdo Dry Valley Lakes, Antarctica - benthic mats | *Phormidesmis priestleyi* ANT.L66.1 | AY493581.1 | 99 | Antarctic lake - benthic microbial mat |
| ***Methylobacterium-1508***[b] | Uncultured bacterium clone LIB079_B_C08 | KM852235.1 | 99 | Biofilm – drinking water | *Methylobacterium brachiatum* strain ZJ0902B96 | KU173699.1 | 98 | East China sea - surface seawaters |
| ***Xenococcaceae-2773***[a] | Uncultured *cyanobacterium* clone FQSS103 | EF522323.1 | 98 | Rocky Mountain, USA - endolithic sandstone community | *Chroococcus* sp. VP2-07 | GQ504036.1 | 91 | Italy and Spain - fountains |
| ***Streptophyta-3354***[b] | Uncultured bacterium clone BJOH-158 | KP724774.1 | 97 | South Shetland Archipelago, Antarctica - glacier ice | *Bryum argenteum* chloroplast | KT343960.1 | 95 | Botany Bay, Antarctica - Antarctic moss |
| ***Leptolyngbya-5202***[a] | Uncultured *cyanobacterium* clone LH16_269 | KM112118.1 | 99 | McMurdo Dry Valley Lakes, Antarctica - benthic mats | *Phormidesmis priestleyi* ANT.L66.1 | AY493581.1 | 99 | Antarctic lake - benthic microbial mat |
| ***Methylobacterium-6342***[b,c] | Uncultured *alphaproteobacterium* clone CN-2_B05 | EF219937.1 | 99 | Antarctica, unvegetated soil environments at Coal Nunatak | *Methylobacterium tardum* strain IHBB 11162 | KR085941.1 | 99 | Suraj Tal, Trans Himalayas - Lake Water |
| ***Rickettsiaceae-7571***[b] | *Candidatus Trichorickettsia mobilis* clone 35 | HG315619.1 | 95 | Wastewater plant | *Candidatus Trichorickettsia mobilis* clone 35 | HG315619.1 | 95 | Wastewater plant |
| ***Pseudomonas-37120***[b] | *Pseudomonas salomonii* strain HS9-MRL | KX128926.1 | 98 | Siachen Glacier, Pakistan - ice, water and sediment | *Pseudomonas antarctica* strain PAMC 27494 | CP015600.1 | 98 | Barton Peninsula, King George Island, Antarctica |
| ***Sediminibacterium-472311***[a] | Uncultured *Sediminibacterium* sp. clone Limnopolar-0.5-B2 | FR848659.1 | 99 | Byers Peninsula, Antarctica - limnopolar lake | *Vibrionimonas magnilacihabitans* strain MU-2 | NR_133888.1 | 99 | Lake Michigan - water sample |

930