ISME

ARTICLE

Microbial connectivity and sorting in a High Arctic watershed

Jérôme Comte[1,2,3,6] · Alexander I. Culley[1,2,4] · Connie Lovejoy[2,3,5] · Warwick F. Vincent[1,3]

Abstract
Aquatic ecosystems in the High Arctic are facing unprecedented changes as a result of global warming effects on the cryosphere. Snow pack is a central feature of northern landscapes, but the snow microbiome and its microbial connectivity to adjacent and downstream habitats have been little explored. To evaluate these aspects, we sampled along a hydrologic continuum at Ward Hunt Lake (latitude 83°N) in the Canadian High Arctic, from snow banks, water tracks in the permafrost catchment, the upper and lower strata of the lake, and the lake outlet and its coastal marine mixing zone. The microbial communities were analyzed by high-throughput sequencing of 16 and 18S rRNA to determine the composition of potentially active Bacteria, Archaea and microbial Eukarya. Each habitat had distinct microbial assemblages, with highest species richness in the subsurface water tracks that connected the melting snow to the lake. However, up to 30% of phylotypes were shared along the hydrologic continuum, showing that many taxa originating from the snow can remain in the active fraction of downstream microbiomes. The results imply that changes in snowfall associated with climate warming will affect microbial community structure throughout all spatially connected habitats within snow-fed polar ecosystems.

Introduction

With ongoing global climate change, projections indicate that the Arctic region will continue to warm more rapidly than the global mean, especially over land ([1] and references therein). Lakes are a major feature of the northern landscape [2, 3], and given their strong dependence on snow, ice and permafrost, Arctic lakes are sentinel

✉ Jérôme Comte
jerome.comte@ete.inrs.ca

1 Centre d'études nordiques (CEN), Université Laval, Québec, QC G1V 0A6, Canada

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

3 Takuvik Joint International Laboratory & Département de biologie, Université Laval, Québec, QC G1V 0A6, Canada

4 Département de biochimie, de microbiologie et de bio-informatique, Université Laval, Québec, QC G1V 0A6, Canada

5 Québec-Océan, Université Laval, Québec, QC G1V 0A6, Canada

6 Present address: Institut national de la recherche scientifique - Centre Eau Terre Environnement, Québec, QC G1K 9A9, Canada

indicators of climate [4, 5]. Climate warming is especially expected to influence snow cover and alter lake ice phenology [4], in turn affecting water flow and the functioning of Arctic lake ecosystems [6].

Snow has wide-ranging effects on ecosystems by modulating temperature, hydrology and climate at local [7] and global [8] scales. The influence of snow on ecosystems is especially marked in the polar regions, given that snowpack covers a vast area of the northern landscape for most of the year. The balance of photosynthesis and respiration in Arctic lakes is linked to the phenology of lake snow cover, which can influence primary production by controlling the availability of light. Snow melt can also influence lake production by controlling nutrient inputs from the surrounding catchment [9]. In addition, snow contains microbes with representatives of the three main domains of life; Bacteria, Archaea, and Eukarya [10–17]. Studies of the biogeography and local adaptions of the snow microbiome [16] and analyses at local and global scales [14] have indicated differences in community structure among regions, but little attention has been given to spatial patterns within connected hydrologic systems.

Recent studies on hydrologic continua have shown that aquatic microbial communities in lakes and streams contain organisms with terrestrial origins [18–20], and that topography and hydrology play an important role in the transfer of microbiota from land to water [21]. Landscape

connectivity is thought to influence microbial community assembly, structure and co-occurrence, and has been investigated in polar soils [22], lakes [18], and rivers [19]. Particular attention has been given to the contributions of glacier melt water to chemistry and microbial assemblages, with evidence of transport of glacier ecotypes downstream [19].

Snow melt is often the main source of water for Arctic lakes feeding into lakes by diffuse overland flows and via conduits of subsurface flow over the permafrost table (i.e., water tracks; [23, 24]). The snowpack acts as a reservoir of water, solutes, and particles, and the resultant meltwater plays a major role in polar landscape connectivity by redistributing matter and energy downstream [25], thereby bridging terrestrial and aquatic ecosystems [26]. These snow-derived conduits discharge water and materials into lakes and provide inorganic nutrients [27], dissolved organic carbon, and contaminants [28]. In addition, they may also be a source of microbial populations to lakes, but this has been little investigated to date.

Ward Hunt Island is located at the northern terrestrial limit of Arctic Canada and is therefore uniquely positioned as an indicator of global change [29]. Pronounced environmental changes in response to warming have been observed and documented at this High Arctic site over the last two decades, including break-up of surrounding ice shelves and the loss of thick perennial ice on Ward Hunt Lake in the centre of the island [30]. Major changes are predicted to continue in the High Arctic; for example, precipitation is likely to increase, but as rainfall rather than snow [31].

Here, we investigated the connectivity of potentially active microbes in assemblages of Bacteria, Archaea, and Eukarya throughout the Ward Hunt Lake watershed. To achieve this, we determined the diversity and composition communities using high throughput Illumina amplicon sequencing targeting 16S (Bacteria and Archaea) and 18S (Eukarya) rRNA along the water continuum from snowbanks via waters tracks and the lake to the coastal mixing zone at the edge of the Arctic Ocean. Specifically, we tested the hypotheses that the Ward Hunt watershed hosts microbial communities that differ among the different aquatic habitats, and that snow represents a source for microbial communities located downstream.

Material and methods

Study sites and sampling

During summer 2014, we sampled a hydrologic continuum within the catchment area of Ward Hunt Lake (83° 05.297'N, 74°09.985'W); a description of the lake and its watershed is given in Paquette et al. [30]. This oligotrophic waterbody lies in a polar desert landscape (mean annual air temperature of around $-17\,°C$) and is covered by thick (>1.5 m) perennial ice during most summers, including 2013, 2014, and 2015 (but with complete loss of lake ice in 2011 and 2012; [30]). At the time of sampling, an ice-free moat had formed on the north-western side of the lake to ~15 m from the shoreline, as shown in Fig. 1b of Mohit et al. [32]. The hydrological continuum starts from snowbanks that supply meltwater into the lake through watertracks: sub-surface, pipe-like conduits over the permafrost table [23]. A small stream at the outlet of the lake connects the inland aquatic environment to the Arctic Ocean (Fig. 1). Six habitats were sampled along this continuum: snowbanks, water tracks, mid-lake epilimnion (surface layer, at 2 m just beneath the ice) and hypolimnion (bottom layer, at 8 m depth), lake outlet and the mixing zone between fresh and marine waters. RNA was collected from each habitat using aseptic techniques and RNAase Zap (Ambion) treated sampling gear to halt RNA degradation. Three separate biological replicates were collected for each of the habitats. For snow sampling, we used a shovel to transfer snow into three sterile 55 oz Whirl-Pack™ sample collection bags (Zeflon). The samples were kept in the dark in a cooler then slowly thawed over 1–2 days at ambient outdoor air temperature resulting in two liters of snowmelt water per snow site. Water tracks were sampled by lifting the overlying biocrusts and collecting the water using 50 ml sterile tubes (Falcon), which were then poured into a 2 l sterile bottle. For the lake, a Jiffy® drill was used to drill through the 2 m-thick lake ice and the water was collected at ca. 2 m (epilimnion) and 8 m (hypolimnion) depth using a Limnos water sampler after vertical profiling of conductivity, temperature, and depth in the water column using a RBR Concerto profiler (data available at Nordicana D database— http://www.cen.ulaval.ca/nordicanad/; [33]). The middle of the lake outlet and outlet water at the edge of land-fast ice that extended 10 m out into the sea from the shoreline (coastal mixing zone) were collected directly by submerging sterile 2 l bottles that were capped when full. All samples were transferred to a field laboratory near Ward Hunt Lake and filtered immediately. Water was sequentially filtered using a peristaltic pump, first through a 20 μm mesh net and a 47-mm diameter, 3 μm pore size polycarbonate filter (Whatman) to remove larger organisms and particles, and then through a 0.2 μm pore-size Sterivex filter (EMD Millipore, Billerica, MA, USA) to collect the microbial cells. The Sterivex filters were preserved with 1.8 ml of RNA*later* (Life Technologies) and stored at $-20\,°C$ until further processing.

Microbial community composition

Microbial RNA was extracted from the Sterivex units using the AllPrep Micro Kit (Qiagen), with DNA and the RNA
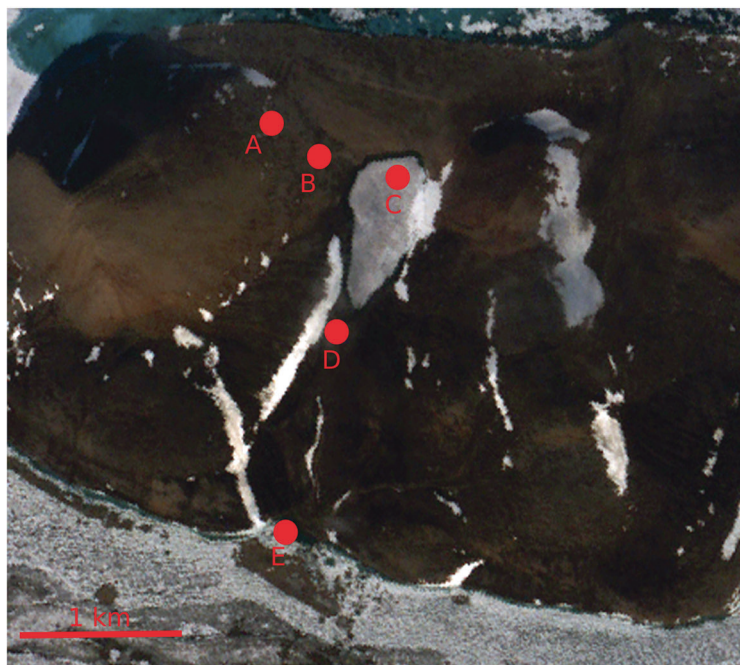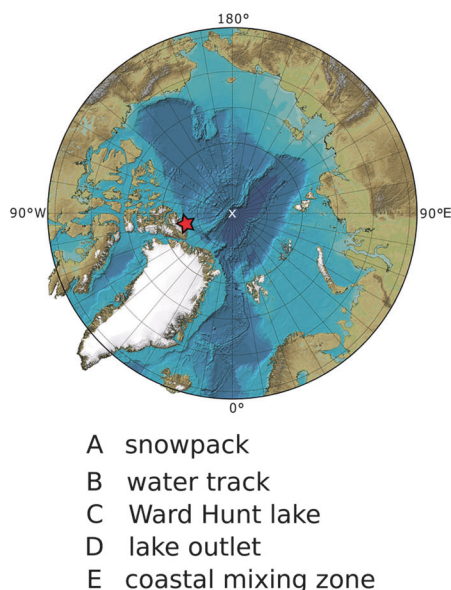
Fig. 1 Left: Location of Ward Hunt Lake (red star) relative to the North Pole (white cross); the base map is the International Bathymetric Chart of the Arctic Ocean (IBCAO) of the Intergovernmental Oceanographic Commission (IOC), version 3.0 released in the public domain on 8 June 2012 [59]. Right: location of sampling sites; background Formosat-2 image courtesy of Pax Arctica and the Planet Action Initiative, © NSPO 2008—National Space Organization, Taiwan, Distribution Spot Image S.A., France, All rights reserved Pointing et al. [60]; Tarnocai et al. [61]

recovered from the same filters as described in Thaler et al. [34]. Following the added DNAase step, RNA purity was verified by a trial PCR as recommended by the manufacturer. RNA was then converted to cDNA using the High Capacity Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions.

The V6−V8 regions of the bacterial and archaeal 16S rRNA and the V4 region of the eukaryotic 18S rRNA which had been converted into cDNA were amplified using the primers described in detail in Comeau et al. [35] but with Illumina adapters. PCR amplification was in two steps. First, 25 μl PCR reactions were performed in separate three template concentrations using non-barcoded primers (PCR1); this was followed by a second set of three PCR reactions using purified PCR product from the first round as template (PCR2). In both cases, each PCR reaction comprised of 0.4 U Phusion high-fidelity DNA polymerase (New England Biolabs), 1 × HF buffer (NEB, UK), 0.5 μM primers, 200 μM deoxynucleotide triphosphates (dNTPs, Invitrogen, USA) mix and between 1 and 5 μL of template cDNA. PCR conditions were as follows for Bacteria: an initial denaturation step at 98 °C for 30 s, 25 (PCR1) or 15 (PCR2) cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 7 min. For Archaea: an initial denaturation step at 98 °C for 30 s, 35 (PCR1) or 15 (PCR2) cycles of 98 °C for 10 s, 52 °C for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 5 min. For Eukaryotes: an initial

denaturation step at 98 °C for 30 s, 30 (PCR1) or 15 (PCR2) cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 4.5 min. Amplicons from PCR1 and PCR2 were purified using magnetic beads (Axygen PCR cleanup kit, Axygen, USA) and quantified spectrophotometrically (Nanodrop, ND-1000).

Library preparation and sequencing on an Illumina MiSeq platform, was done at the Plateforme d'Analyses Génomiques, Institut de Biologie Intégrative et des Systèmes, Université Laval (Québec, Canada). The raw reads have been pooled with other sequences from microbial mats collected in the same area [32] and deposited in the NCBI sequence read archive (SRA) under the study accession SRP078933: records SRX301419 to SRX3014244.

Sequence processing

All sequence data processing was carried out using the UPARSE pipeline [36]. Initial preprocessing included removing reads <300 nucleotides and the remaining reads were quality controlled and chimeras detected using UCHIME and the Ribosomal Database Project (RDP) training set v.14. The cleaned reads were then clustered into operational taxonomic units (OTUs) using a 99% sequence identity cutoff. OTU sequence representatives were aligned using PyNAST [37] with the pre-aligned Greengenes 16S core set [38] for Archaea and Bacteria and the Silva

pre-aligned data set for Eukaryotes as a template. OTU taxonomic assignment was performed using the RDP classifier [39] with a 50% assignment confidence cutoff and the SILVA v.111 as a reference database [40]. Non-aligned sequences, Metazoa and Fungi in the eukaryote dataset, reads classified as chloroplasts or Archaea in the Bacteria data, and Bacteria in the Archaea data were removed from the analyses.

Phylogenetic analyses

Phylogenetic analyses were based on a dendrogram constructed with an approximate maximum-likelihood (ML) approach using the 'accurate mode' in FastTree v.2.1 [41]. Weighted and unweighted distances [42] among the different microbial communities were all computed based on the ML tree.

Faith's phylogenetic diversity [43], was determined using R package 'picante' v1.5 [44]. Community phylogenetic structure was further explored based on the net relatedness index (NRI) as defined by Webb and coworkers [45] at the level of the whole community and at the genus level of *Polaromonas*, which was used as an example taxon for population metrics. Briefly, positive NRI values indicate phylogenetic clustering in which the community is composed of taxa that are more evolutionarily related (environmental filtering of taxa that share specific traits) than in randomly assembled communities. In contrast, negative NRI values suggest phylogenetic overdispersion, in which the taxa are less related than expected by chance, for example, due to competition.

Statistics

All analyses were performed on subsampled datasets of equal sequencing depths (27,025 sequences per sample for Bacteria, 1162 sequences per sample for Archaea, and 35,327 sequences per sample for Eukarya), which corresponded to the lowest number of reads retrieved from the respective domains. Patterns in beta-diversity among the different habitats were visualized using principal coordinate analyses (PCoA) and their significance tested by permutation MANOVA (Permanova in R). All statistical analyses were performed in R 3.0.3 (R [46]) and charts drawn using ggplot2 [47]. The assumptions of normality and homoscedasticity for an analysis of variance were confirmed using a Shapiro–Wilk test. ANOVA and an a posteriori TukeyHSD test (seed OTUs, NRI) or Kruskal–Wallis followed by the Dunn [48] post-hoc test (RNA and DNA concentrations) were used to investigate for differences among the different habitats.

The algorithm SourceTracker [49] was used to identify the most probable source of dispersal for each local community within the watershed and more specifically to estimate the contribution of snow as a source of dispersal to the community of the WHI aquatic ecosystem. Samples from the epilimnion and hypolimnion of WHL were highly similar based on UniFrac distances, and the corresponding sequences were therefore merged and identified as the "lake community". Each local community (e.g. snow, lake) was set either as a source or a sink depending on the direction of the evaluation; potential contribution as a source or community potential origin. For example, snow was set as the source community and independent SourceTracker analyses (four in total) were carried out for each sink (receiving habitats; i.e., water tracks, lake, lake outlet and coastal mixing zone). Analyses were conducted for each domain of life on subsampled datasets as described above. We additionally evaluated the potential of snow to be a sink with the downstream habitats being a source.

The relative importance of snow was further investigated at the population level. To do this, we first assessed whether an OTU was rare using the 0.1% threshold as suggested by Pedrós-Alio [50]. We then plotted the proportion of sequences for OTUs detected in the different downstream habitats sampled in this watershed as a function of their corresponding proportion in the snow, using the criteria that an OTU should be detected at least once in the snow and in one of the downstream habitats. In total, 2248 OTUs met this criterion. An OTU home habitat was assigned based on the one in which it had the highest proportion. In this analysis, we focused on the identification of snow 'seed OTUs' and their contribution to WHI downstream habitats using the procedure described in Ruiz-González et al. [20]. Snow 'seed OTUs' were defined as the OTUs that had a lower relative abundance in snow in comparison to all downstream habitats. To test whether the rise in proportional rRNA abundance of certain 'seed taxa' from the snowbank to downstream environments was the result of a decrease in total community size, we tested if samples from downstream habitats had lower concentrations of RNA and DNA than upstream habitats.

The main isolation source (habitat from which an OTU sequence has previously been detected) of bacterial, archaeal, and eukaryotic OTUs was identified using the Seqenv algorithm [51], where representative sequences of the OTUs were submitted to a BLASTn analysis against the NCBI nt database and a 99% identity threshold to determine earlier source records for each taxon.

Results

After cleaning and quality processing, we retained 1,051,165 reads for Bacteria, 239,594 reads for Archaea and 1,803,099 reads for Eukarya, which at the 99% similarity level clustered as 3911, 419 and 4751 OTUs, respectively.

Table 1 Species richness and phylogenetic characteristics of the communities retrieved from the different habitats

	Sequences	OTUs	Unique sequences	Unique OTUs	PD	NRI
Bacteria						
Snow	$99,607 \pm 11,922^a$	1655 ± 127^{bc}	492 ± 41^a	44 ± 2^c	43.6 ± 3.7^b	9.8 ± 0.4
Water track	$28,412 \pm 2173^d$	2124 ± 68^a	826 ± 248^a	255 ± 41^a	70.7 ± 4.5^a	2.5 ± 1.9
Lake epilimnion	$60,939 \pm 10,758^b$	1474 ± 11^{cd}	13 ± 7^c	8 ± 3^d	33.9 ± 1.8^c	4.9 ± 3.1
Lake Hypolimnion	$69,515 \pm 12,897^{ab}$	1254 ± 119^d	24 ± 9^{bc}	12 ± 4^d	28.8 ± 3.4^c	9.5 ± 1.0
Lake outlet	$54,398 \pm 11,065^{bc}$	1939 ± 186^{ab}	68 ± 24^b	33 ± 12^c	51.7 ± 4.9^b	6.2 ± 4.3
Mixing zone	$37,518 \pm 6809^{cd}$	1901 ± 220^{ab}	993 ± 445^a	115 ± 20^b	53.6 ± 4.1^b	2.0 ± 4.6
Archaea						
Snow*	$25,198 \pm 33,579$	66 ± 55	11950	25^*	2.1 ± 0.8^b	10.3 ± 3.0^a
Water track	$22,767 \pm 16,910$	156 ± 71	1252 ± 1056	32 ± 17	4.9 ± 1.7^a	5.1 ± 2.5^a
Lake epilimnion	2368 ± 881	109 ± 4	1	1^*	5.4 ± 0.4^a	6.0 ± 0.9^a
Lake hypolimnion	$33,879 \pm 21,726$	139 ± 3	188 ± 162	11 ± 4	5.1 ± 0.1^a	11.1 ± 1.0^a
Lake outlet	3143 ± 288	164 ± 8	22 ± 1	1	6.6 ± 0.3^a	5.5 ± 1.7^a
Mixing zone	1957 ± 1195	158 ± 22	1001 ± 731 ns	20 ± 1	6.9 ± 1.1^a	-4.0 ± 1.3^b
Eukarya						
Snow	$115,265 \pm 13,371^a$	674 ± 248^c	2713 ± 1226^a	48 ± 14^b	28.8 ± 10.4^b	6.4 ± 1.2^{abc}
Water track	$102,305 \pm 12,003^{ab}$	2015 ± 101^{ab}	11197 ± 6091^a	527 ± 57^a	88.8 ± 4.1^a	14.3 ± 2.6^a
Lake epilimnion	$117,516 \pm 17103^a$	1068 ± 60^{bc}	27 ± 31^b	6 ± 4^c	33.0 ± 2.4^b	8.0 ± 1.1^{ab}
Lake hypolimnion	$57,891 \pm 36,181^b$	755 ± 301^c	129 ± 120^b	11 ± 3^c	26.0 ± 5.2^b	2.8 ± 4.2^{bc}
Lake outlet	$101,738 \pm 23,232^{ab}$	1880 ± 123^{ab}	172 ± 92^b	33 ± 11^b	69.7 ± 5.2^a	9.0 ± 0.7^{ab}
Mixing zone	$106,318 \pm 28,586^{ab}$	2046 ± 84^a	17647 ± 4962^a	466 ± 130^a	73.7 ± 3.7^a	1.4 ± 0.9^c

Values are averages for triplicates (±SD). PD and NRI were calculated from the subsampled datasets. Letters are for Tukey posthoc tests (different letters indicate statistical differences at $p < 0.05$)

*Results from only one replicate available

Species richness (number of OTUs) of Bacteria and Eukarya varied greatly along the hydrologic continuum, with higher values in the water tracks (Table 1). Species richness of Archaea was considerably lower than for Bacteria and Eukarya, and did not vary systematically along the hydrological continuum. Analyses of the phylogenetic structure of the prokaryotic (Bacteria and Archaea) and eukaryotic communities at the OTU level showed positive NRI values (Table 1), indicating that the majority of the communities in the watershed clustered phylogenetically: the co-occurring taxa were more related to each other than would be expected by chance. Only Archaea originating from the oceanic mixing zone, showed significant negative NRI values. This suggests that these archaeal groups showed less relatedness than expected by chance (i.e., the taxa were phylogenetically overdispersed).

PCoA on unweighted UniFrac distances showed that the different habitats had distinct microbial communities (Fig. 2a), and a similar pattern was observed using weighted UniFrac distances (SI Fig. 1). Permanova analyses further confirmed significant differences among the habitats for both weighted (Bacteria: $P = 0.001$, $R^2 = 0.77$; Archaea: $P = 0.001$, $R^2 = 0.79$; Eukarya: $P = 0.001$, $R^2 = 0.78$) and unweighted UniFrac distances (Bacteria: $P = 0.001$, $R^2 = 0.90$; Archaea: $P = 0.001$, $R^2 = 0.84$; Eukarya: $P = 0.001$, $R^2 = 0.89$). As shown in Fig. 2a, there was a clear separation of water track, snow, lake outlet and coastal mixing zone communities. The two lake communities (hypolimnion and epilimnion) tended to cluster more closely together.

Although there were clear differences in community composition among habitats, the different habitats shared a variable number of OTUs (Fig. 2d–f) with a high proportion for Bacteria (up to 30% of the total number of OTUs), but only 3% of Archaea OTUs and 5% of Eukarya OTUs shared among all habitats. Among all the habitats, the water tracks included the highest level of unique OTUs (15%, 11%, and 18% of total OTUs for Bacteria, Archaea, and Eukarya, respectively).

Betaproteobacteria dominated the composition of bacterial communities in all the habitats sampled (Fig. 3a), especially the genus *Polaromonas*. However, we found differences in the phylogenetic structure, as measured by NRI of *Polaromonas* representatives, among habitats (SI Fig. 2). Some bacterial groups showed a degree of habitat specificity with Acidobacteria, Fibrobacteriales, and Deltaproteobacteria mostly detected in the water track (Fig. 3a).
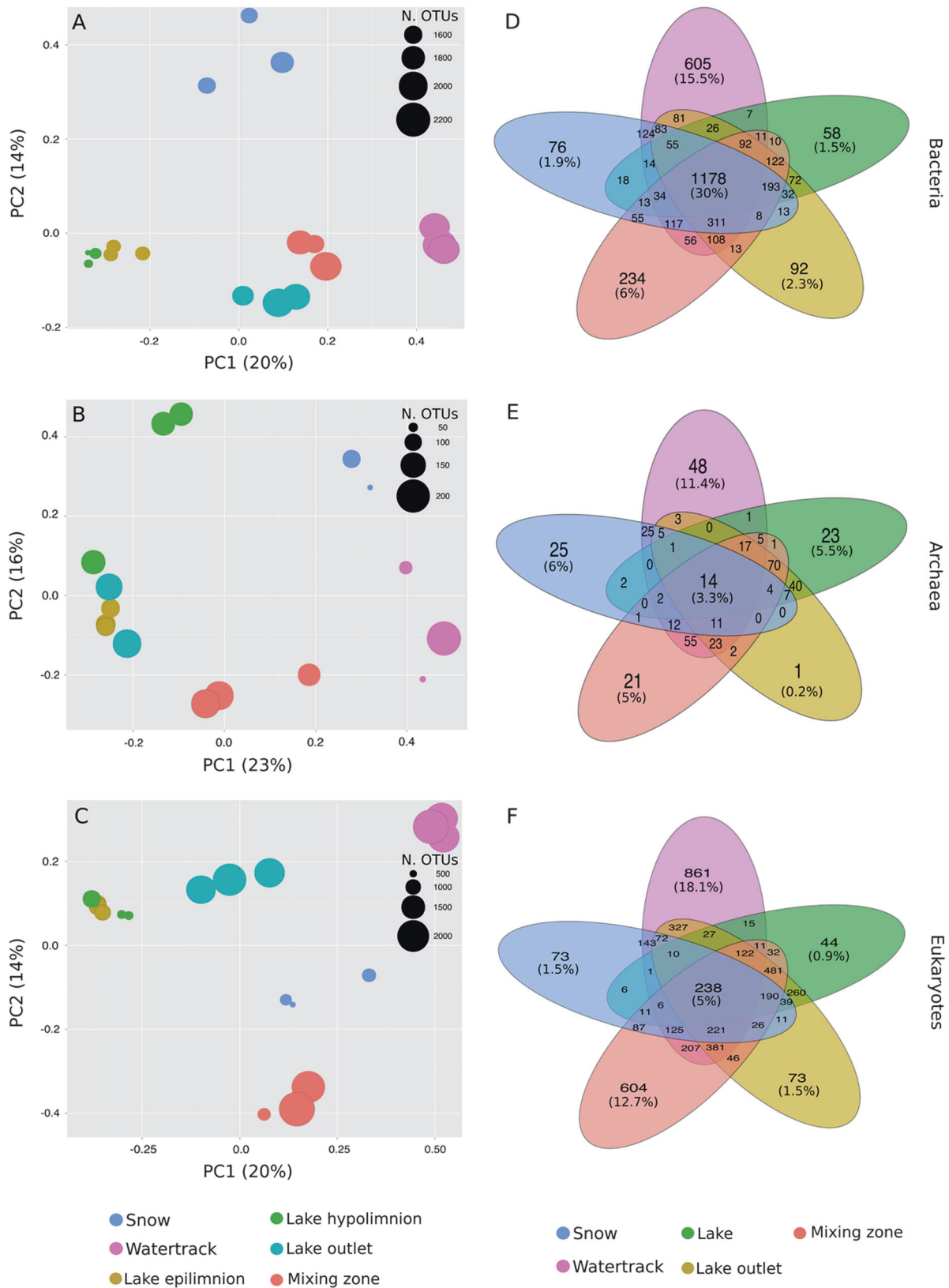
Fig. 2 Principal coordinate analyses based on Unifrac unweighted distances among (**a**) bacterial, (**b**) archaeal and (**c**) eukaryotic community samples. Venn diagrams showing the number and proportion of shared and unique OTUs among the different habitats for (**d**) bacterial, (**e**) archaeal and (**f**) eukaryotic community samples

Bacteroidetes gradually increased along the hydrologic continuum going from 3% of reads in the snow to up to 31% in the coastal mixing zone. Alphaproteobacteria were mostly detected in this zone and were rarer in the other habitats.

Archaeal communities changed substantially along the continuum, with a dominance of Thaumarchaeota in the snow and Euryarchaeota in downstream habitats (Fig. 3b). The relative dominance of taxa within Euryarchaeota varied by site with water track communities more diverse and evenly distributed with members identified as Halobacteriales, Thermoplasmatales, while the lake and stream communities were largely dominated by Halobacteriales, especially in the epilimnion, Methanomicrobiales in the hypolimnion and to a lesser extent to Methasarcinales (17% of reads detected in the epilimnion). Communities from the coastal mixing zone were mainly composed of Halobacteriales and marine Thaumarchaeota.

The dominant eukaryotic taxa in all habitats were in the Cercozoa and Chrysophyceae (Fig. 3c); the latter represented up to 70% and 46% of the total 18S sequences detected in the snow and coastal mixing zone, respectively. Ciliates showed some degree of habitat preference, with higher proportions in the water tracks and coastal mixing zone. Dinoflagellate OTUs dominated lake-habitats (epilimnion, hypolimnion, and the lake outlet, with 61.7%, 28.2%, and 47.1% of total number of reads, respectively), whereas they represented a lower proportion of the coastal mixing zone community and less than 1% of the reads in the snow and water tracks.

The SourceTracker algorithm was applied to identify sources of microbial taxa in each local community, and the results showed that internal turnover, which is the contribution of hypolimnion community to epilimnion assemblages and vice versa, explained the majority of the community composition identified in Ward Hunt Lake (SI Fig. 3). However, the water tracks and the coastal mixing zone sites showed a larger fraction of unknown sources suggesting that other components not sampled here, for example, soil and Arctic Ocean microbiota, respectively, would have made non-trivial contributions.

The program Seqenv was used to match our sequences with public records and to identify their habitat provenance. Most of the bacterial and eukaryotic OTUs were previously detected in freshwater ecosystems, mainly in lakes, whereas the Archaea appeared associated with soil sequences. In addition, the results showed that some of these OTUs had isolation sources identified as part of the cryosphere with OTUs associated with permafrost, glaciers, fjords, high elevations, and sea ice. These cryosphere sources accounted for 8% of for Bacteria, 2% for Archaea, and 17% for Eukarya, notably 15% of Eukarya were associated with sea ice. The snow bacterial OTUs represented up to 14% of the

lake community. In contrast, less than 2% of archaeal and eukaryotic OTUs identified in downstream habitats had a snow origin (Table 2).

We further used SourceTracker to examine to what extent each downstream habitat could represent a source to the snow microbiota. The results showed almost all of the most probable sources for snow community composition were identified as 'unknown' for Eukarya and Archaea. For Bacteria, the lake outlet and to a lesser extent water track and the coastal zone were the most probable sources for the snow community, while the lake community did not appear to be a source for the snow microbiota (Table 2, SI Fig. 3).

To further assess the contribution of snow OTUs to the community composition of the downstream habitats, we first compared the proportion of sequences shared between the snow and the other habitats. Based on the criterion that a 'shared OTU' was detected at least once in the snow and in one of the downstream habitats, this indicated 2248 Bacteria, 84 Archaea, and 1187 Eukarya OTUs (Fig. 4) were shared. Hence, Fig. 4 showed the distribution of these 'shared OTUs' among the different habitats and was used to identify "seed OTUs" (see below). The results further indicated that the most abundant sequences for bacterial and eukaryotic OTUs detected in downstream habitats were in low proportion in the snow (upper left panel in Fig. 4). Although some dominant bacterial taxa in downstream habitats were also abundant in the snow (upper right panel in Fig. 4), this was not the case for either the Archaeal or Eukarya.

We identified 366, 24, and 163 by seed 'OTUs' in upstream snow for Bacteria, Archaea, and Eukarya respectively (Fig. 5a–c). The highest abundance of sequences for these snow 'seed OTUs' in downstream habitats was in the lake. The Betaproteobacteria (in particular members of the family Comamonadaceae), the archaeal group Methanomicrobiales and dinoflagellates (Eukarya) comprised the majority of the snow taxa found in downstream habitats (Fig. 5d–f, SI Table 1). As a relative indication of standing and potentially active biomass, we note that snow had on average the lowest concentration of both DNA and RNA in comparison to the other habitats, with no significant differences among the other habitats (SI Fig. 4). No significant differences were detected in the RNA:DNA ratio among the different habitats (Kruskal–Wallis chi-squared $= 6.7$, df $= 5$, $P = 0.25$).

Discussion

The High Arctic snowbanks sampled in this study contained diverse microbiota in all three domains, and its community composition was distinct relative to communities downstream (Fig. 2). Previous work on snow microbiology at this

Fig. 3 Taxonomic composition of Ward Hunt Island microbial communities based on the relative abundance of (**a**) bacterial, (**b**) archaeal 16S, and (**c**) eukaryotic 18S sequences
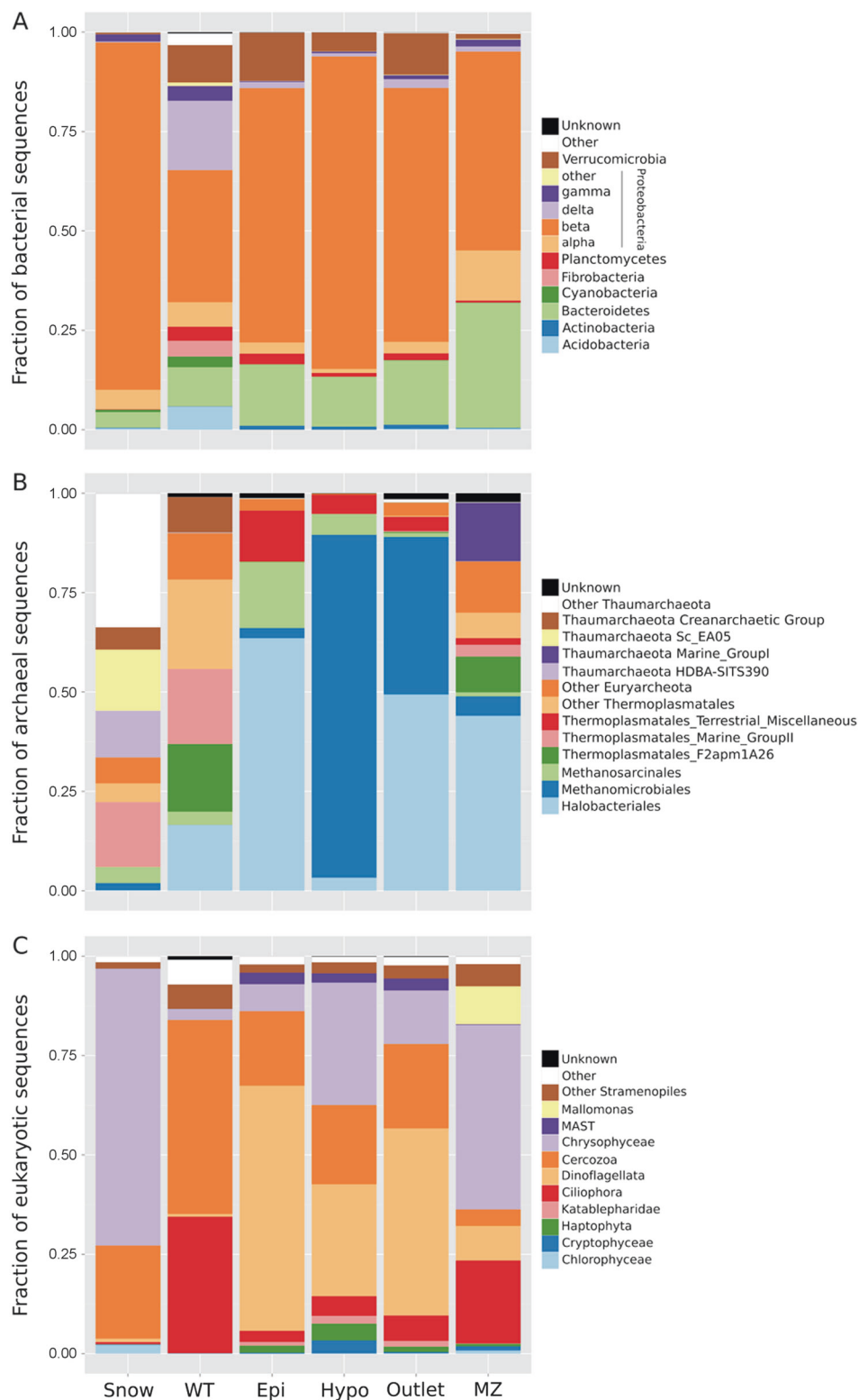
site was based on clone library analysis of rDNA [11], and showed the presence of airborne taxa from diverse sources, including *Polaribacter*, *Loktanella* (marine), and Acidobacteria (terrestrial), as recorded here. The high-throughput analyses in the present study provided a much more detailed depiction of the snow microflora, and based on rRNA (cDNA) rather than genomic DNA, the results gave information on taxa that are more likely to be active in this habitat [52]. We acknowledge that this approach does not allow investigating other important ecological

Table 2 Snow contribution as a source or sink to local community composition

	WT	Lk	Out	MZ
Source				
Bacteria	4.2 (2.3–5.9)	14.3 (9–21)	4.3 (3.1–6.6)	18.1 (16.7–19.6)
Archaea	2.2 (0.9–3.2)	0.03 (0.01–0.06)	0.2 (0.1–0.2)	1.2 (0.8–1.6)
Eukaryota	0.01 (0.001–0.02)	0.25 (0.1–0.3)	0.6 (0.4–0.9)	1.0 (0.3–2.1)
Sink				
Bacteria	11.7 (3.8–20.6)	1.5 (1–2)	69.4 (55–85.4)	11.8 (6.1–15.5)
Archaea	(0.4–1.21)[a]	(0–1.5)[a]	(0–0.12)[a]	(0.06–0.09)[a]
Eukaryota	0.04 (0.001–0.06)	0.37 (0.003–1.5)	0.17 (0.02–0.32)	0.04 (0.008–0.09)

Values represent average proportion (%, min – max) of downstream communities with snow identified as source (SourceTracker analyses). WT, Lk, Out and MZ refer to water track, lake, lake outlet, and marine-freshwater mixing zone, respectively
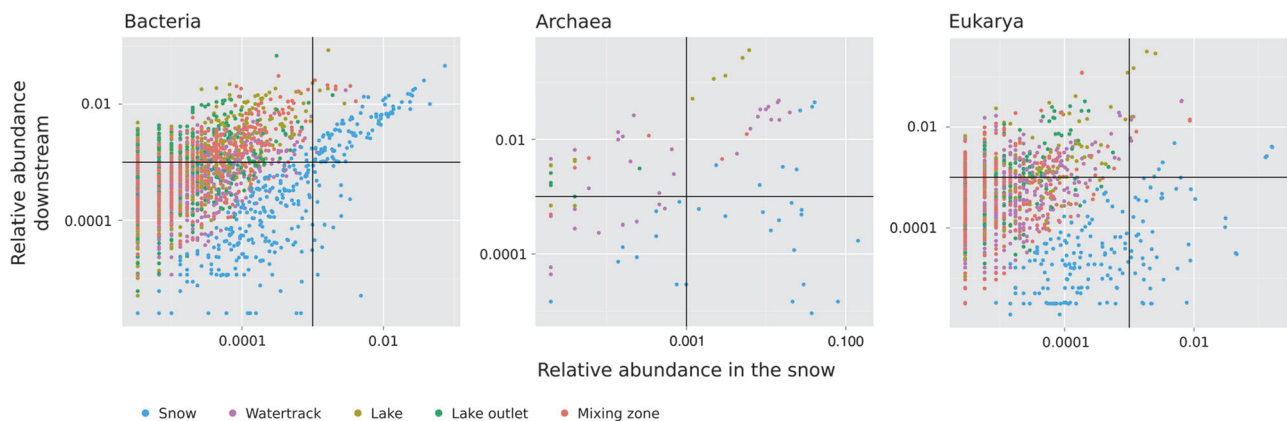
[a]Only two replicates

Fig. 4 Relative abundance of sequences for bacterial, archaeal, and eukaryotic OTUs that appear at least once in the snow and in at least one downstream habitat (2248 OTUs in total). An OTU was assigned to one of the habitats based on its maximal relative abundance, which was calculated for pooled triplicate samples from each habitat

questions. For example, the connectivity patterns presented here are only for the potential active fraction of the community but it is likely that the connectivity among the different habitats may be stronger as many of the dispersed organisms may be in a dormant state and therefore contributing to the local seedbank of habitat located downstream. Importantly, we cannot say whether the dominant OTUs identified in this study were also the most abundant ones from the whole DNA pool or were rare, functionally important taxa (as in [53]). For Bacteria, the dominance by Betaproteobacteria is in line with previous studies in snow [54], but the dominant eukaryotes were Chrysophytes, not Chlorophytes as found elsewhere and characteristic of red or green snow [54, 55].

The Ward Hunt Island snow community included chemolithotrophic bacteria involved in iron cycling (e.g. *Acidiphilium*), as well as bacterial methylotrophs such as *Methylotenera* and microbes that utilize C1 compounds including methane, such as *Methylacidiphilum*, Archaeal methanogens (*Methanomicrobiales*) were also well represented.

The fact that few taxa were unique to a given type of habitat (Table 1), together with the clear habitat-specific segregation of communities (Fig. 2a–c), points to a strong local selection of species (species sorting) from the shared pool of microbes at the watershed scale or metacommunity (Fig. 2a–c), as indicated by the analysis of the phylogenetic structure of these communities. Of particular note, of the shared pool of species across habitats, most of the OTUs that reached their highest sequence proportion in downstream habitats had low sequence abundance in the snow (Fig. 4).

The rise in proportional rRNA abundance of certain 'seed taxa' from the snowbank to downstream environments implies one or a combination of three possible scenarios: (a) the cells have grown to become more proportionately abundant in the downstream communities; (b) the number of ribosomes per cell has increased downstream in these taxa (as in ref. [53]); and (c) the total community size has deceased, with other taxa dropping their rRNA contribution to a much greater extent than the 'seed taxa'. Scenario (c) seems least likely, because the total
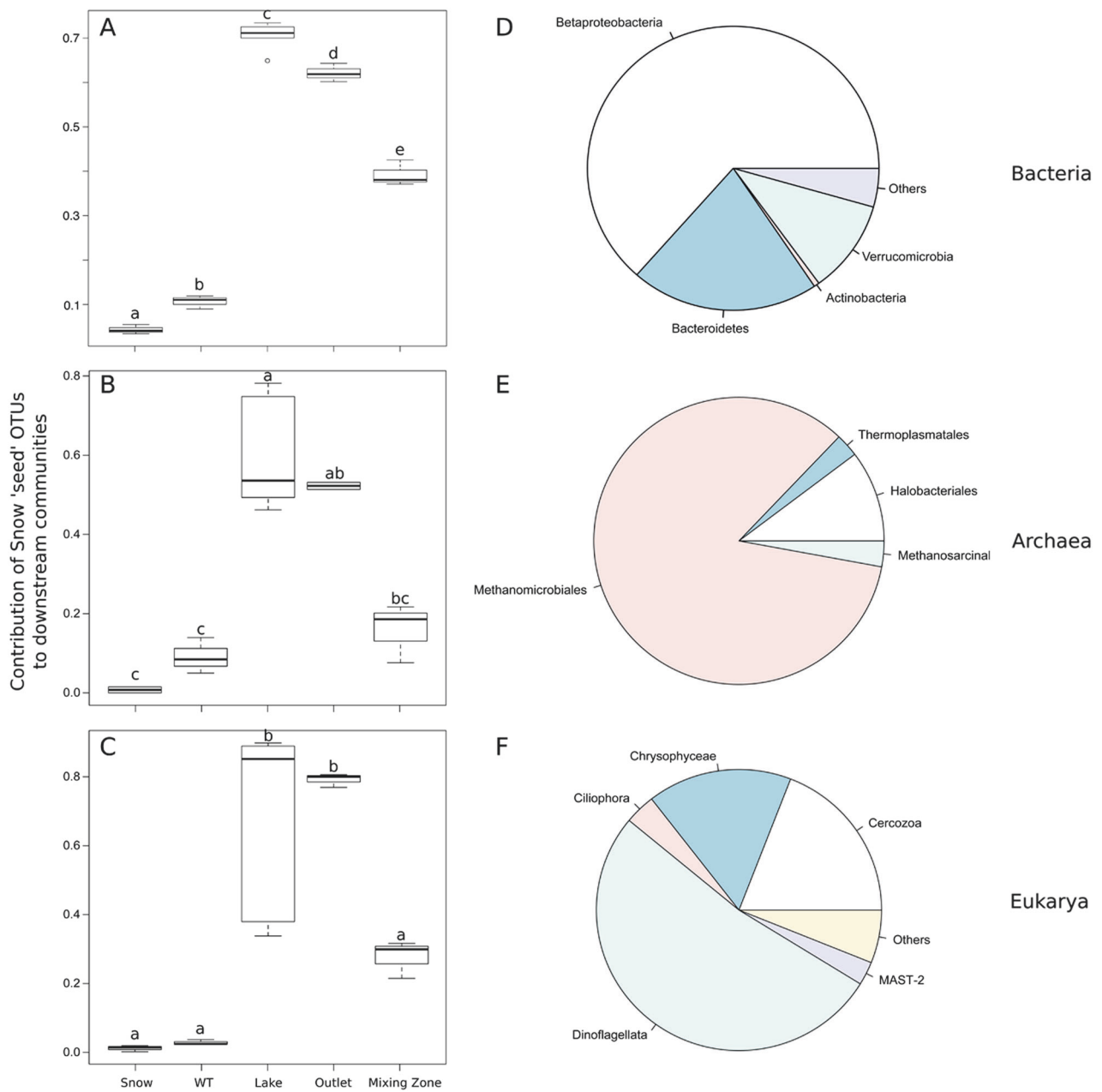
Fig. 5 Contribution of snow seed operational taxonomic units ('seed OTUs') to total sequences (with 1 representing 100% of total sequences) in the downstream habitats (panels **a–c**). Letters indicate significant differences in the contribution of snow OTUs among habitats. Panels **d–f** showed the taxonomic composition of the snow seed pools for Bacteria, Archaea, and Eukaryotes, respectively

RNA concentration was higher not lower downstream (SI Fig. 4), and there was no significant difference in RNA:DNA ratios among sites. Both (a) and (b) indicate that certain snow biota ('seed taxa') rise in ribosomal proportion throughout the hydrological continuum, suggesting increased potential activity.

The detectable signature of the snow microbial assemblages all the way from the water tracks to the coastal mixing zone with the Arctic Ocean (Fig. 5) indicates that snow could be a microbial source for downstream habitats including the lake. Our results indicate that the contrary is unlikely, and that dispersal is limited from the small moat region of open water on Ward Hunt Lake to the snowpack. This is in accordance with previous studies conducted in Arctic and boreal hydrologic networks [18–20] and supports the notion that lake ecosystems can be seeded by organisms that originate from their surrounding catchments. The presence of terrestrial and marine biota in the snow microbiome is consistent with the aerial deposition of materials from the nearby open ocean and exposed soils.

In this study, we were specifically interested in the dispersal of cryophilic organisms from the snowpack to the lake. *Polaromonas* was particularly abundant across the watershed. However, the phylogenetic clustering in the snow and the coastal mixing zone habitats in particular suggests that the downstream habitats are composed of a diverse community of *Polaromonas* ecotypes. In particular, the snow and mixing zone communities appeared to be composed of *Polaromonas* ecotypes that were more phylogenetically related to each other than those retrieved in the lake. This in turn suggests that environmental forcing in the snow selected for ecotypes that shared common and most likely specific traits in regards to ambient environmental conditions. In contrast, the lake seems to host more diverse *Polaromonas* representatives, which suggests a greater diversity of potential niches in the lake as opposed to snow. This is in line with previous studies reporting a high ratio of local to regional diversity in *Polaromonas* [56], which provides opportunities to exploit distinct metabolic niches [57, 58]. Our results imply that the Arctic lake microbiota is partly composed of snow organisms, which in the context of global warming has important implications for ecosystem functioning.

The water tracks contained a striking diversity of taxa, consistent with previous analyses based on pyrosequencing [24]. In addition to their high diversity, the water tracks had other distinct features compared to the other habitats. Firstly, the number of unique taxa was high; although similar for the Archaea, the water track bacterial and eukaryotic communities had 6 and 11 times, respectively, more unique taxa than the snowbanks (Table 1). Secondly, the phylogenetic diversity was also extremely high: 1.6 (Bacteria), 2.3 (Archaea), and 3.1 (Eukarya) times the values in the snowbanks (Table 1). Finally, these communities were also the most phylogenetically clustered (highest NRI values). This is consistent with the organic-rich nature of the subsurface environment that the water tracks flow through [23]; although this flow is rapid, the transit time appears sufficiently long for the water to generate a distinct microbial signature. The abrupt decline in diversity in the lake, and the loss of so many taxa unique to water tracks, suggest that the water track community is specialized towards the subsurface soil environment with many species that are unable to survive in oligotrophic freshwaters (SI Fig. 5).

The only habitat with similarly high values to the water tracks was the coastal mixing zone, with entrainment of marine taxa from the Arctic Ocean. These communities were the least clustered (Table 1). In fact, NRI values for Archaea were negative, which suggests that members from these communities were less ecologically similar to each other than they would be by chance (overdispersion across the phylogeny), again reflecting the mixing of communities from disparate sources.

Snow cover is expected to change in the future, and models predict that future precipitation in the Arctic may be predominantly in the form of rain including in the winter where intense rainfall is expected to occur intermittently [31]. This will strongly impact the hydrology of the High Arctic, including a reduced period of snow cover, more rapid snow cover melt and increased surface runoff over shorter time periods (e.g., [26] and references therein), which in turn, will have consequences for freshwater habitats including the water tracks and Ward Hunt Lake. Under these future conditions, the sources of potential contributors to the diversity of the lake microbiota are likely to change in character, for example, via changes in snow accumulation and melt patterns or increased rainwater runoff and permafrost thaw. The water tracks and their highly diverse communities may be especially vulnerable to the effects of liquid water, since their geomorphology and flow regime depend on the permafrost table and snowbank water source, both of which are likely to change substantially with ongoing climate warming.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Barry RG. The Arctic cryosphere in the twenty-first century. Geogr Rev. 2016;107:69–88.
2. Paltan H, Dash J, Edwards M. A refined mapping of Arctic lakes using Landsat imagery. Int J Remote Sens. 2015;36:5970–82.
3. Verpoorter C, Kutser T, Seekell DA, Tranvik LJ. A global inventory of lakes based on high-resolution satellite imagery. Geophys Res Lett. 2014;41:6396–402.
4. Mueller DR, Van Hove P, Antoniades D, Jeffries MO, Vincent WF. High Arctic lakes as sentinel ecosystems: cascading regime shifts in climate, ice cover, and mixing. Limnol Oceanogr 2009; 54:2371–85.
5. Williamson CE, Saros JE, Vincent WF, Smol JP. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. Limnol Oceanogr. 2009;54:2273–82.
6. Wrona FJ, Johansson M, Culp JM, Jenkins A, Mård J, Myers-Smith IH, et al. Transitions in Arctic ccosystems: ecological implications

of a changing hydrological regime. J Geophys Res Biogeosci. 2016;121:650–74.

7. Callaghan TV, Johansson M, Brown R, Groisman P, Labba N, Radionov V, et al. The changing face of Arctic snow cover: a synthesis of observed and projected changes. Ambio. 2011;40:17–31.

8. Vavrus S. The role of terrestrial snow cover in the climate system. Clim Dyn. 2007;29:73–88.

9. Vincent WF, Callaghan TV, Dahl-Jensen D, Johansson M, Kovacs KM, Michel C, et al. Ecological implications of changes in the Arctic cryosphere. Ambio. 2011;40:87–99.

10. Christner BC, Mosley-Thompson E, Thompson LG, Zagorodnov V, Sandman K, Reeve JN. Recovery and identification of viable bacteria immured in glacial ice. Icarus 2000;144:479–85.

11. Harding T, Jungblut AD, Lovejoy C, Vincent WF. Microbes in high arctic snow and implications for the cold biosphere. Appl Environ Microbiol. 2011;77:3234–43.

12. Hell K, Edwards A, Zarsky J, Podmirseg SM, Girdwood S, Pachebat JA, et al. The dynamic bacterial communities of a melting High Arctic glacier snowpack. ISME J. 2013;7:1814–26.

13. Lopatina A, Medvedeva S, Shmakov S, Logacheva MD, Krylenkov V, Severinov K. Metagenomic analysis of bacterial communities of Antarctic surface snow. Front Microbiol. 2016;7:398.

14. Lutz S, Anesio AM, Raiswell R, Edwards A, Newton RJ, Gill F, et al. The biogeography of red snow microbiomes and their role in melting arctic glaciers. Nat Commun. 2016;7:11968.

15. Wunderlin T, Ferrari B, Power M. Global and local-scale variation in bacterial community structure of snow from the Swiss and Australian Alps. FEMS Microbiol Ecol 2016;92:fiw132.

16. Cameron KA, Hagedorn B, Dieser M, Christner BC, Choquette K, Sletten R, et al. Diversity and potential sources of microbiota associated with snow on western portions of the Greenland Ice Sheet. Environ Microbiol. 2015;17:594–609.

17. Cameron KA, Stibal M, Hawkings JR, Mikkelsen AB, Telling J, Kohler TJ, et al. Meltwater export of prokaryotic cells from the Greenland ice sheet. Environ Microbiol. 2017;19:524–34.

18. Crump BC, Amaral-Zettler LA, Kling GW. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME J. 2012;6:1629–39.

19. Hauptmann AL, Markussen TN, Stibal M, Olsen NS, Elberling B, Baelum J, et al. Upstream freshwater and terrestrial sources are differentially reflected in the bacterial community structure along a small Arctic river and its estuary. Front Microbiol. 2016;7:1474.

20. Ruiz-González C, Niño-García JP, Del Giorgio PA. Terrestrial origin of bacterial communities in complex boreal freshwater networks. Ecol Lett. 2015;18:1198–206.

21. Niño-García JP, Ruiz-González C, del Giorgio PA. Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. ISME J 2016;10:1755–66.

22. Ferrari BC, Bissett A, Snape I, van Dorst J, Palmer AS, Ji M, et al. Geological connectivity drives microbial community structure and connectivity in polar, terrestrial ecosystems. Environ Microbiol. 2015;18:1834–49.

23. Paquette M, Fortier D, Vincent WF, Water tracks in the High Arctic: a hydrological network dominated by rapid subsurface flow through patterned ground. Arct Sci. 2017;3:334–53.

24. Steven B, Lionard M, Kuske CR, Vincent WF. High bacterial diversity of biological soil crusts in water tracks over permafrost in the high arctic polar desert. PLoS ONE. 2013;8:e71489.

25. Peters DPC, Groffman PM, Nadelhoffer KJ, Grimm NB, Collins SL, Michener WK, et al. Living in an increasingly connected world: a framework for continental-scale environmental science. Front Ecol Environ. 2008;6:229–37.

26. Bring A, Fedorova I, Dibike Y, Hinzman L, Mård J, Mernild SH, et al. Arctic terrestrial hydrology: a synthesis of processes, regional effects and research challenges. J Geophys Res Biogeosci. 2016;121:621-649.

27. Jones HG, Pomeroy JW. Early spring snowmelt in a small boreal forest watershed: influence of concrete frost on the hydrology and chemical composition of streamwaters during rain-on-snow events. 2001;58:209–18.

28. Veillette J, Muir DCG, Antoniades D, Small JM, Spencer C, Loewen TN, et al. Perfluorinated chemicals in meromictic lakes on the northern coast of Ellesmere Island, High Arctic, Canada. Arctic. 2012;65:245–56.

29. Vincent WF, Fortier D, Lévesque E, Boulanger-Lapointe N, Tremblay B, Sarrazin D, et al. Extreme ecosystems and geosystems in the Canadian High Arctic: Ward Hunt Island and vicinity. Ecoscience. 2011;18:236–61.

30. Paquette M, Fortier D, Mueller DR, Sarrazin D, Vincent WF. Rapid disappearance of perennial ice on Canada's most northern lake. Geophys Res Lett 2015;42:1433–40.

31. Bintanja R, Andry O. Towards a rain-dominated Arctic. Nat Clim Change. 2017;7:263–7.

32. Mohit V, Culley A, Lovejoy C, Bouchard F, Vincent WF. Hidden biofilms in a far northern lake and implications for the changing Arctic. npj Biofilms Micro. 2017;3:17.

33. NEIGE. Water column physico-chemical profiles of lakes and fiords along the northern coastline of Ellesmere Island, v. 1.1 (1954-2016). Nordicana 2017;D27, https://doi.org/10.5885/45445CE-7B8194DB81754841.

34. Thaler M, Vincent WF, Lionard M, Hamilton AK, Lovejoy C. Microbial community structure and interannual change in the last epishelf lake ecosystem in the north polar region. Front Microbiol. 2017;3:275.

35. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS ONE. 2011;6:e27492.

36. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10:996–8.

37. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266–7.

38. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72.

39. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.

40. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–96.

41. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5:e9490.

42. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71:8228–35.

43. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Cons. 1992;61:1–10.

44. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26:1463–4.

45. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. Phylogenies and community ecology. Annu Rev Ecol Syst. 2002;33:475–505.

46. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. http://www.R-project.org/

47. Wickham H. ggplot2: elegant graphics for data analysis. New York, NY: Springer; 2009.

48. Dinno A. dunn.test: Dunn's test of multiple comparisons using rank sums. R package version 1.3.5. 2017. https://CRAN.R-project.org/package=dunn.test. Accessed on June 8, 2018.

49. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8:761–3.

50. Pedrós-Alio C. The rare bacterial biosphere. Annu Rev Mar Sci. 2012;4:449–66.

51. Sinclair L, Ijaz UZ, Jensen LJ, Coolen MJL, Gubry-Rangin C, Chroňáková A, et al. Seqenv: linking sequences to environments through text mining. PeerJ. 2016;4:e2690.

52. Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. 2013;7:2061–8.

53. Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA, Quince C, Battin TJ. Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. Environ Microbiol. 2014;16:2514–24.

54. Lutz S, Anesio AM, Edwards A, Benning LG. Linking microbial diversity and functionality of arctic glacial surface habitats. Environ Microbiol. 2017;19:551–65.

55. Anesio AM, Lutz S, Chrismas NAM, Benning LG. The microbiome of glaciers and ice sheets. npj Biofilms Micro. 2017;3:10.

56. Darcy JL, Lynch RC, King AJ, Robeson MS, Schmidt SK. Global distribution of *Polaromonas* phylotypes—evidence for a highly successful dispersal capacity. PLoS ONE. 2011;6:e23742.

57. Mattes TE, Alexander AK, Richardson PM, Munk AC, Han CS, Stothard P, et al. The genome of *Polaromonas* sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. Appl Environ Microbiol. 2008;74:6405–16.

58. Yagi JM, Sims D, Brettin T, Bruce D, Madsen EL. The genome of *Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. Environ Microbiol. 2009;11:2253–70.

59. Jakobsson M, Mayer L, Coakley B, Dowdeswell JA, Forbes S, Fridman B, et al. The International Bathymetric Chart of the Arctic Ocean (IBCAO) version 3.0. Geophys Res Lett. 2012;39:L12609.

60. Pointing SB, Burkhard B, Convey P, Gillman LN, Körner C, Leuzinger S, et al. Biogeography of photoautotrophs in the high polar biome. Front Plant Sci. 2015;6:692.

61. Tarnocai C, Canadell JG, Schuur EAG, Kuhry P, Mazhitova G, Zimov S. Soil organic carbon pools in the northern circumpolar permafrost region. Glob Biogeochem Cycles. 2009;23:BG2023.