



# Warming Climate Is Reducing the Diversity of Dominant Microbes in the Largest High Arctic Lake

Graham A. Colby<sup>1</sup>, Matti O. Ruuskanen<sup>1</sup>, Kyra A. St.Pierre<sup>2</sup>, Vincent L. St.Louis<sup>2</sup>, Alexandre J. Poulain<sup>1</sup> and Stéphane Aris-Brosou<sup>1,3\*</sup>

<sup>1</sup> Department of Biology, University of Ottawa, Ottawa, ON, Canada, <sup>2</sup> Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada, <sup>3</sup> Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada

Temperatures in the Arctic are expected to increase dramatically over the next century, and transform high latitude watersheds. However, little is known about how microbial communities and their underlying metabolic processes will be affected by these environmental changes in freshwater sedimentary systems. To address this knowledge gap, we analyzed sediments from Lake Hazen, NU Canada. Here, we exploit the spatial heterogeneity created by varying runoff regimes across the watershed of this uniquely large high-latitude lake to test how a transition from low to high runoff, used as one proxy for climate change, affects the community structure and functional potential of dominant microbes. Based on metagenomic analyses of lake sediments along these spatial gradients, we show that increasing runoff leads to a decrease in taxonomic and functional diversity of sediment microbes. Our findings are likely to apply to other, smaller, glacierized watersheds typical of polar or high latitude ecosystems; we can predict that such changes will have far reaching consequences on these ecosystems by affecting nutrient biogeochemical cycling, the direction and magnitude of which are yet to be determined.

## OPEN ACCESS

### Edited by:

Anne D. Jungblut,  
Natural History Museum,  
United Kingdom

### Reviewed by:

Jérôme Comte,  
Université du Québec, Canada  
Nico Salmaso,  
Fondazione Edmund Mach, Italy

### \*Correspondence:

Stéphane Aris-Brosou  
sarisbro@uottawa.ca

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 11 May 2020

**Accepted:** 28 August 2020

**Published:** 07 October 2020

### Citation:

Colby GA, Ruuskanen MO, St.Pierre KA, St.Louis VL, Poulain AJ and Aris-Brosou S (2020) Warming Climate Is Reducing the Diversity of Dominant Microbes in the Largest High Arctic Lake. *Front. Microbiol.* 11:561194. doi: 10.3389/fmicb.2020.561194

**Keywords:** high arctic, microbial ecology, metagenome assembled genomes (MAGs), high-throughput sequencing, climate change

## INTRODUCTION

Climate change is amplified in polar regions, where near-surface temperatures have increased almost twice as fast as elsewhere on Earth over the last decade (Overpeck et al., 1997; Serreze and Francis, 2006; Screen and Simmonds, 2010). Climate models predict that annual surface temperatures will increase in the Arctic by as much as 8°C by 2100 relative to the 1981–2010 average (Mingle, 2020). These changes are already having dramatic consequences on physical (Bliss et al., 2014; O'Reilly et al., 2015; Laudon et al., 2017), biogeochemical (Frey and McClelland, 2009; Lehnher et al., 2018), and ecological (Smol et al., 2005; Wrona et al., 2016) processes across Arctic ecosystems. During the winter months (January–March) of 2016 and 2018 surface temperature in the central Arctic were already 6°C warmer (Mingle, 2020). While we are starting to understand the effect of thawing permafrost on microbial communities and shallow aquatic ecosystems (McCalley et al., 2014; Crevecoeur et al., 2015; Hultman et al., 2015; Mackelprang et al., 2016), our knowledge of how microbial communities in large aquatic ecosystems at high latitudes respond to environmental changes is comparatively lacking in part due to the large spatial variability to be expected in such systems. Furthermore, lakes are broadly considered sentinels of

climate change, as they integrate physical, chemical, and biological changes happening through their watersheds (Williamson et al., 2009); however, their microbial community structure and function are relatively understudied, in particular in the Arctic.

To date, much of the research performed on microbial communities in Arctic lakes has been limited to studies that were mostly based on partial 16S rRNA gene sequencing (Stoeva et al., 2014; Mohit et al., 2017; Thaler et al., 2017; Ruuskanen et al., 2018; Cavaco et al., 2019). While these studies are useful to understand the structure of these microbial communities, they provide limited functional insights and can be biased as they often rely on sequence databases where environmental microbes, specifically from the Arctic, may be underrepresented (Ruuskanen et al., 2018, 2019). More critically, being circumscribed both in space and in time, previous studies only offer snapshots of microbial communities and hence, have a limited power to predict how microbial communities might respond to climate change.

To predict the effect of climate change on microbial functional diversity in Arctic lake sediments, we focused on Lake Hazen, the world's largest High Arctic lake (82° N, 71° W) by volume which is estimated at  $5.14 \times 10^{10} \text{ m}^3$  (Köck et al., 2012). Lake Hazen is ultra-oligotrophic, and its physical, chemical, and biological limnology has been previously described in depth in St. Pierre et al. (2019b). In this work, we exploited two important properties of Lake Hazen. First, its watershed is already experiencing the effects of climate change, as increasing regional temperatures are leading to more glacial melt, permafrost thaw, and increased runoff from the watershed into the lake in warmer years relative to cooler ones (Lehnherr et al., 2018). Second, its tributaries are highly heterogeneous, fed by 11 glaciers ranging from 6 to 1,041 km<sup>2</sup> in surface area, and annual runoff volume approximately scaling with their size (from <0.001 to 0.080 km<sup>3</sup> in 2016; St. Pierre et al., 2019b). Glacial meltwaters feed the lake from late June through to the end of August.

It is this spatial heterogeneity in runoff that we used to evaluate the possible consequences of climate change on microbial structure and functional diversity in High Arctic sediment, acknowledging that the consequences of increasing temperature are likely more plural and complex. Although this approach simplifies the effects of climate change, glacial runoff provides the main source of sediment and nutrients to this extremely oligotrophic environment (Lehnherr et al., 2018). Therefore, increasing temperature and subsequent runoff are assumed to be the primary drivers of alterations to this microbial ecosystem. To this effect, we sampled lake sediments along two transects representing low (L transect: samples L1 [shallow] and L2 [deep]) and high (H: samples H1 [shallow] and H2 [deep]) seasonal runoff volume, as well as at a single site that received negligible runoff (C site; Figure 1A). In order to assess the potential level of connectivity between lake sediments and upstream soil (Crump et al., 2012; Comte et al., 2018; Hermans et al., 2020), we also collected soil samples (S sites) from three sites in the dried streambeds of the tributaries, on the northern shore between the two transects to assess soil influence on microbial communities present in the sediments.

We then leveraged untargeted metagenomics analyses to draw an inventory of dominant microbes, assumed to be the most critical to nutrient cycling and the most relevant to the dynamics of microbial communities. These reconstructed Metagenome Assembled Genomes (MAGs) (Bowers et al., 2017) allowed us to assess the quantitative impact of a change of runoff regime, from low to high, on both the structure of sediment microbial communities and their functional potential. We show that an increase in runoff volume and resultant sedimentation rates, as predicted under climate change scenarios for the region, could lead to a reduced diversity of the dominant microbial community and of their functional potential.

## RESULTS

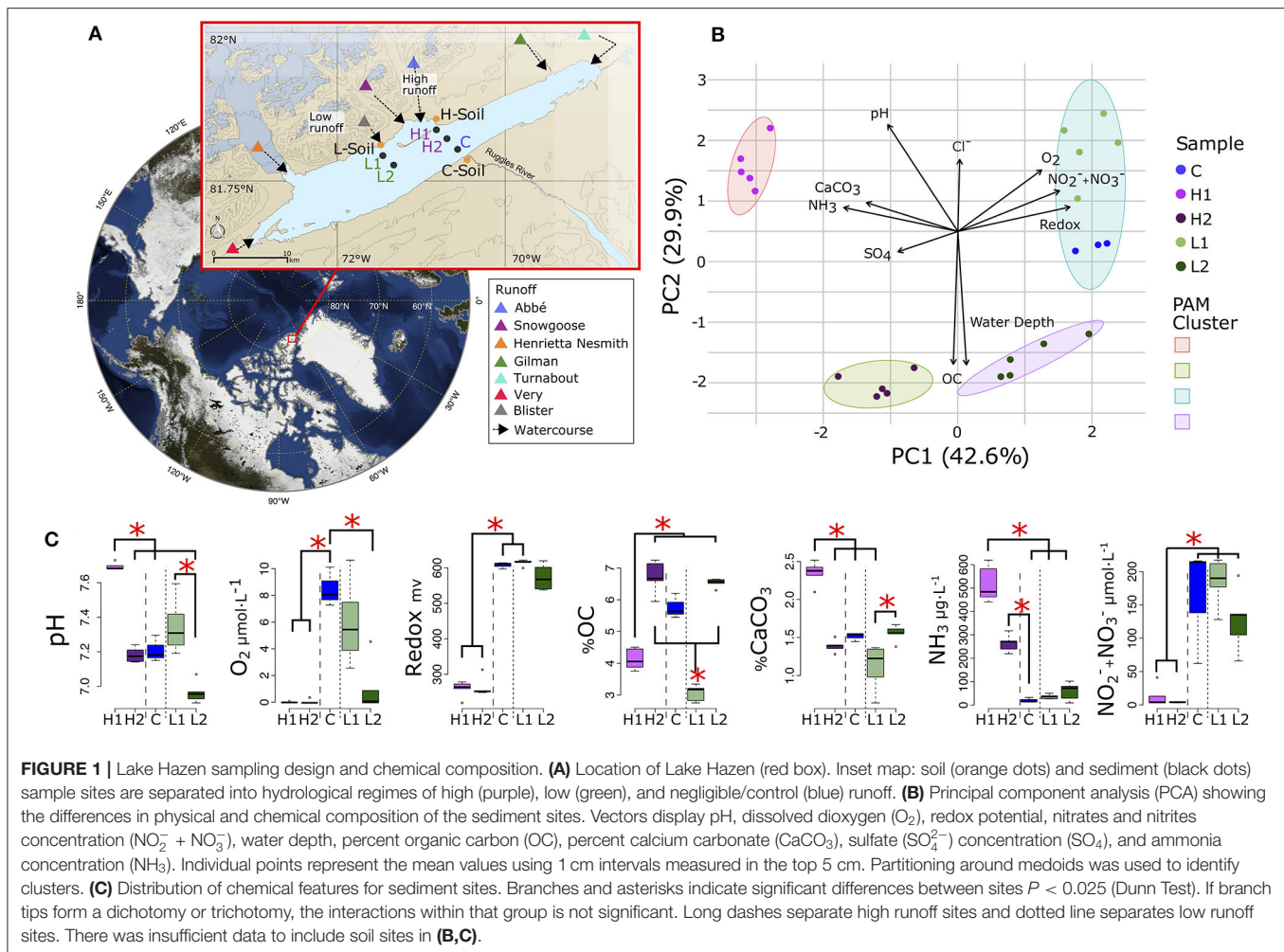
### Characterization of the Physical and Geochemical Environments

We first characterized how geochemical properties of the sediments varied along and between the two transects (soil chemistry could not be obtained due to the limited number of cores that could be collected and shipped back, and to in-lab restrictions on chemical measurements) by re-analyzing data from St. Pierre et al. (2019b). Sediment samples from these five sites clustered into four distinct geochemical groups (Figure 1B) that reflect spatial variability in glacial runoff, the primary hydrological input to the lake. Indeed, PC1 explained 43% of the total variance, and differentiated the L and high H runoff transects, while PC2 (29.9%) separated each transect according to their depth.

Along PC1, higher concentrations of ammonia (NH<sub>3</sub>) and sulfate (SO<sub>4</sub><sup>2-</sup>) in the porewaters, and a greater percentage of calcium carbonate in the sediments, were present in the H transect. However, higher concentrations of dioxygen (O<sub>2</sub>), nitrates/nitrites (NO<sub>3</sub><sup>-</sup>/NO<sub>2</sub><sup>-</sup>), and greater redox potential were present in the L transect and the control (C) sites. Along PC2, sediment organic carbon (OC), and porewater pH and Cl<sup>-</sup>, were more determinant when discriminating between the shallow (L1 and H1) and deep (L2 and H2) sites of both transects. Rather than grouping spatially with the H transect, the C sites were most chemically similar to L1 (Figure 1C, Supplementary Figure 1). The shallow sites were not significantly different from each other in pH measurements or OC concentrations, but were both significantly different from the deeper sites suggesting that although most chemical features were similar within each transect, some features might still be influenced by their spatial proximity to the shoreline or depth of the overlying water column (Figure 1C).

### Contrasting Low vs. High Runoff Transects Revealed a Decrease in Biodiversity

With such a clear geochemical separation of the transects along PC1 (43% of explained variance) and significant spatial contrasts (Figure 1C), we had the right context to evaluate the influence of runoff gradients on sediment microbial diversity. We assembled a total of 300 (290 bacterial and 10 archaeal) MAGs that were >50% complete and with <10% contamination (Minimum

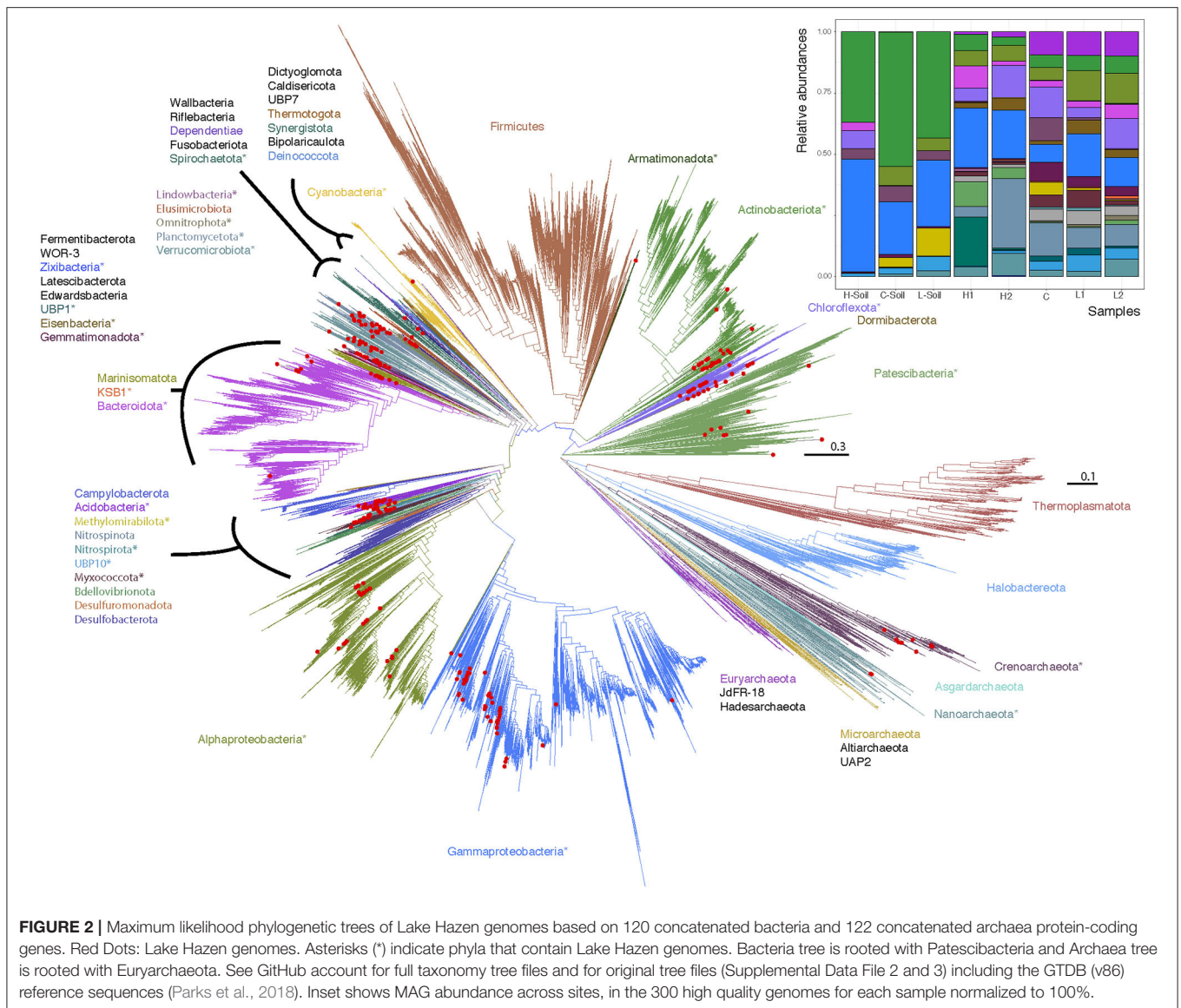


information about a metagenome-assembled genome [MIMAG] guidelines; Bowers et al., 2017, **Supplementary Tables 1, 2**). By constructing phylogenetic trees for Bacteria and Archaea, we noted that while most major phyla were represented in the MAGs, no Firmicutes and only a small number of Archaea were identified (**Figure 2**). In contrast, Gammaproteobacteria ( $n = 50$ ), Actinobacteria ( $n = 31$ ), Alphaproteobacteria ( $n = 24$ ), Chloroflexoata ( $n = 30$ ), Planctomycetota ( $n = 24$ ), and Acidobacteriota ( $n = 19$ ) were the most commonly recovered taxa across the entire watershed. Uncultured phyla comprised  $\sim 11\%$  of reconstructed MAGs, including representatives from multiple taxa: Eisenbacteria ( $n = 12$ ), Patascibacteria ( $n = 9$ ), Omnitrophica ( $n = 5$ ), KSB1 ( $n = 1$ ), Armatimonadota ( $n = 1$ ), Lindowbacteria ( $n = 1$ ), USBP1 ( $n = 1$ ), UBP10 ( $n = 1$ ), and Zixibacteria ( $n = 1$ ). We note that missing sequence data in the gene alignments had a minimal impact on the trees (**Supplementary Text, Supplementary Figures 2, 3**).

However, these MAGs were not evenly distributed across all sites (**Figure 2**, inset; **Supplementary Figure 4**). To quantify this uneven distribution, we determined the site where each genome was most abundant. Based solely on this information, we performed an unsupervised clustering (*t*-SNE), and found

that the directions defined by sediment-laden water flowing from the shallow to the deep site within each transect in the projection space were almost orthogonal between transects (see arrows in **Figure 3**, showing the relative positions of the hyperplanes defined by water flow)—which suggests that, as in a Principal Component Analysis, these directions are independent (although a *t*-SNE cannot strictly speaking be interpreted in this way), and hence that transitioning from the L to H transect leads to a dramatic shift in microbial communities. This shift can also be characterized from a diversity point of view, where both PCoA (**Supplementary Figure 5**) and a DPCoA (**Supplementary Figure 6**) results suggest differences between the L and H transects—at least from a qualitative standpoint.

To quantify the direction and the significance of these structural and diversity shifts, we focused on the phylum level (to extract large-scale patterns), and calculated the relative proportions of each of the reconstructed 300 MAGs at each site, and tallied these numbers by phylum, over the 25 phyla represented in our data. We did this along each transect—essentially pooling sites H1/H2 together to represent the H transect, and doing the same for sites L1/L2 (the L transect), while keeping proportions for the S and C sites separate. Note that sites



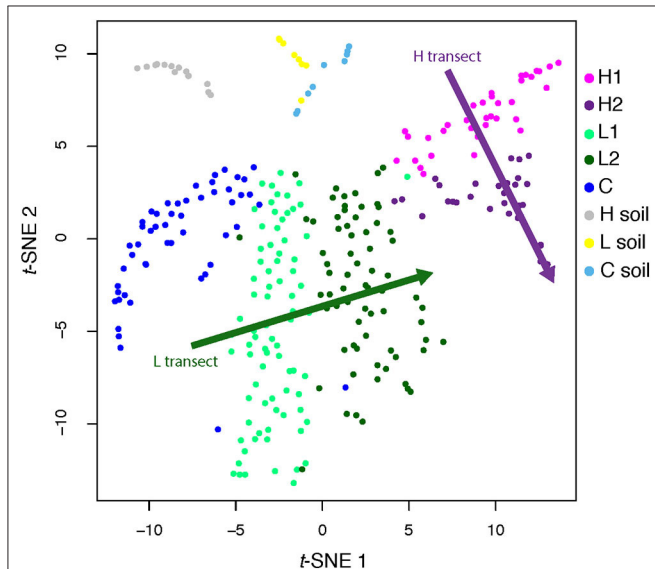
S and C were included at this stage to gauge the potential level of connectivity between lake sediments and both upstream and downstream soils. Hierarchical clustering on this table of MAGs proportions by phyla vs. sites showed a divergence from the L to H transects [following the (((L,C),H),S) clustering pattern; **Figure 4A**, inset—see also **Supplementary Figures 7, 8**], with the C site (negligible runoff) grouping with the L sites, the S sites (soils) with the H sites, and confirming the clear contrast between the two transects in terms of taxa proportions (see **Figure 3**). To test if these taxa proportions tended to increase or decrease when transitioning from L to H along the (((L,C),H),S) clustering pattern, we fitted linear models (ANOVA) regressing the proportions of each of the 25 phyla against sites, ordered as per their hierarchical clustering (L→C→H→S). Essentially, we regressed a single data point for each of the four classes (L, C, H, and S), so that *P*-values could not be obtained, but slope

could be estimated (**Figure 4A**). Strikingly, most of these slopes were negative (22 out of 25; binomial test:  $P = 7.8 \times 10^{-8}$ ; see also **Supplementary Figure 8A**), demonstrating a significant decrease in diversity at the phylum level as one goes from low to high runoff regimes. Because the inclusion of the C and S sites in these analyses could add noise when specifically testing for a change in the proportion of phyla during a L to H transition, we reran these analyses without the C and S sites. In spite of some phyla changing slope sign, suggesting a certain instability in our results due to small sample size, we found that most of the estimated slopes were negative (19 out of 25; binomial test:  $P = 0.0073$ ; see also **Supplementary Figure 8B**), still providing evidence of a significant decrease in diversity at the phylum level as one goes from low to high runoff regimes.

An NMDS ordination allowed us to detect the geochemical features associated with this shift in microbial communities

(Supplementary Figure 9). In the sediments,  $\text{NH}_3$  concentrations ( $P = 0.03$ ),  $\text{NO}_2^- / \text{NO}_3^-$  concentrations ( $P = 0.03$ ), and redox potential ( $P = 0.03$ ) were significant in determining the distribution of MAGs (permutation test:  $P < 0.05$ ). We further observed that the sites with the greatest diversity (L/C sites) were also those with the greatest redox potential, and  $\text{O}_2$  and  $\text{NO}_3^- / \text{NO}_2^-$  concentrations. Sites with

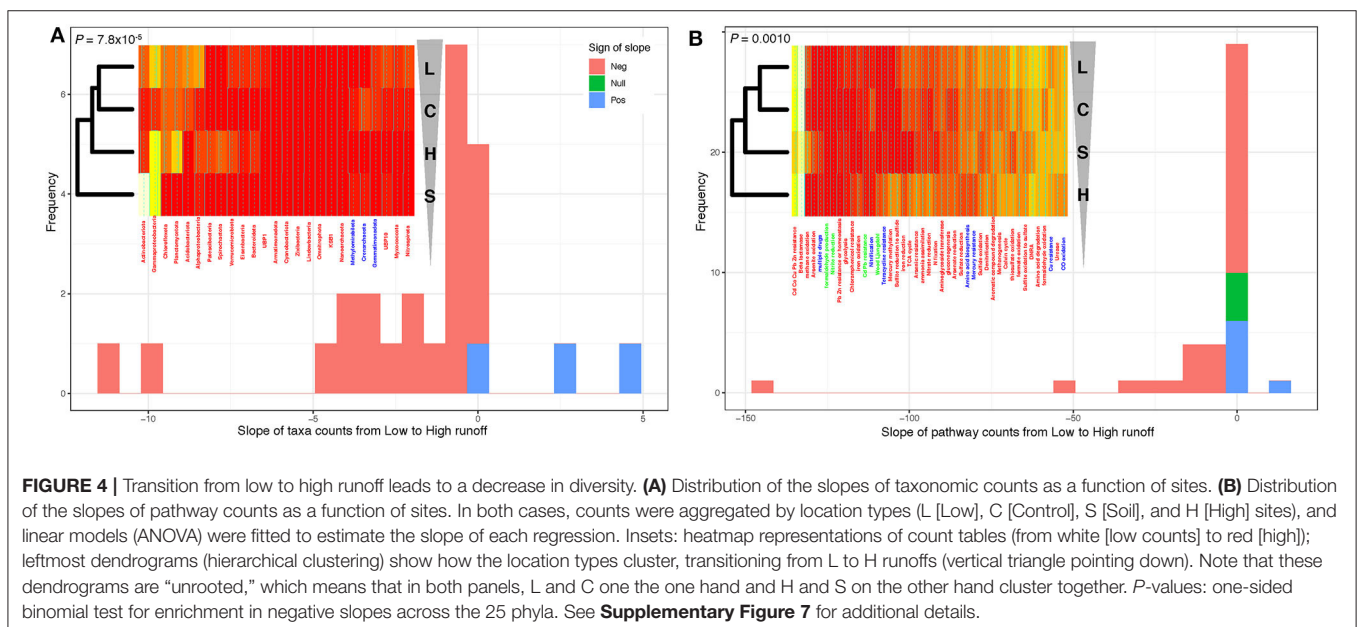
the lowest microbial diversity (H sites), contained greater  $\text{NH}_3$  and  $\text{SO}_4^{2-}$  concentrations, and lower redox potential relative to the C and L sites. In addition to gradients shaped by the interplay between microbial metabolism and local geochemical constraints, the physical disturbances associated with high sedimentation rates also likely contributed to the loss of microbial diversity; however, we cannot quantify the relative importance of each of these processes here.



**FIGURE 3** | t-SNE analysis of genome abundance for each sediment sample. Each of the 300 shown genomes was assigned to the sample where it has the greatest abundance. Shaded arrows display the approximate direction of water flow, from upstream to downstream, for the high (green) and low (purple) transects. Note that these arrows do not indicate any kind of convergence.

## Contrasting Low vs. High Runoff Transects Also Revealed a Loss of Functional Potential

To assess the functional implications of this decrease of biodiversity, we assigned metabolic functions and pathways to proteins in each MAG. We focused on genes and pathways involved in key elements, targeting carbon, nitrogen, and sulfur cycling (Supplementary Figures 10, 11). Only the most abundant genomes per site were reported within each phylum (Supplementary Figure 12), allowing us to compute the proportions of functions and pathways in each of the 25 phyla present in reconstructed MAGs across the hydrological regimes (Supplementary Figures 13, 14). Their hierarchical clustering (Supplementary Figure 13) led to a picture consistent with the ones derived from both geochemical (Figure 1) and taxonomic abundances (Figure 4A). Indeed, the two transects were again clearly separated [clustering pattern ((L,C),S),H); Figure 4B, inset], and fitting linear models regressing function/pathway proportions against sites showed that, again, most of these slopes were negative (binomial test:  $P = 0.0010$ ). Forcing the same site ordering as for the taxonomic abundances (L→C→H→S as in Figure 4A, inset) led to similar results (binomial test:  $P = 7.8 \times 10^{-5}$ ), demonstrating a significant decrease in metabolic diversity when going from the L to the H transect.



**FIGURE 4** | Transition from low to high runoff leads to a decrease in diversity. **(A)** Distribution of the slopes of taxonomic counts as a function of sites. **(B)** Distribution of the slopes of pathway counts as a function of sites. In both cases, counts were aggregated by location types (L [Low], C [Control], S [Soil], and H [High] sites), and linear models (ANOVA) were fitted to estimate the slope of each regression. Insets: heatmap representations of count tables (from white [low counts] to red [high]); leftmost dendrograms (hierarchical clustering) show how the location types cluster, transitioning from L to H runoffs (vertical triangle pointing down). Note that these dendrograms are “unrooted,” which means that in both panels, L and C one the one hand and H and S on the other hand cluster together.  $P$ -values: one-sided binomial test for enrichment in negative slopes across the 25 phyla. See Supplementary Figure 7 for additional details.

More specifically, we found that marker genes whose product is implicated in carbon and sulfur metabolisms significantly decreased when going from the L to H, while nitrogen metabolism was unaffected (**Supplementary Table 3**, see **Supplementary Text** for details). When considering the individual functions present or absent across the transects, we noted that most oxidative pathways (CO, methane, formaldehyde, sulfide, sulfite) appeared less common in the H transect (**Supplementary Figure 10**), corresponding to lower oxygen concentrations and constraints on aerobic metabolism. Furthermore, while most carbon fixation processes were shared between the two transects, carbon oxidation and reduction reactions regulated through Wood-Ljungdahl pathway were only observed in the H transect, where sedimentary conditions were anoxic throughout the first 5 cm (see Figure 4 in St.Pierre et al., 2019b), consistent with a more reductive environment.

## DISCUSSION

Even if Arctic microbial communities are changing rapidly (Hultman et al., 2015), there is still a dearth of long-term time series observations. As a first step toward addressing this point, we used Lake Hazen's spatial geochemical heterogeneity to evaluate the structural and functional response of lake sediment microbial communities to varying runoff, already shown to increase in this warming High Arctic environment (Lehnher et al., 2018). Such an approach can reasonably be interpreted from the lens of a space-for-time design, which assumes that spatial and temporal variations are not only equivalent (Blois et al., 2013; Lester et al., 2014), but also stationary (Damgaard, 2019). Whether this latter condition is met cannot be known, but in the absence of any time-series documenting the effect of climate change on lake sediment microbial communities in the High Arctic, the space-for-time design becomes a convenience, if not a necessity (Pickett, 1989).

Using metagenomics along two transects experiencing heterogeneous runoff conditions, we presented evidence that climate change, as it drives increasing runoff and sediment loading to glacial lakes, will likely lead to a decrease in both diversity and functional potential of the dominant microbial communities residing in lake sediments. Note that we specifically focused here on the dominant microbes, that is those for which we could reconstruct the MAGs, in order to (i) have a phylogenetic placement of the corresponding organisms based on a large number of marker genes (**Figure 2**), rather than partial 16S rRNA gene sequences as usually done (Ruuskanen et al., 2018), and (ii) be able to predict almost complete functional pathways for each of these organisms to test the impact of a change of runoff (**Figure 4**), rather than inferring function from taxonomic affiliation (Ruuskanen et al., 2018). On the other hand, by focusing on dominant microbes, we lose taxonomic depth, and may miss key ecological and geochemical roles played by the "rare biosphere" (Lynch and Neufeld, 2015), and its associated shifts in the context of a warming climate.

Such a decrease in taxonomic and functional diversity may not be unique to Lake Hazen, where rising temperatures have resulted

in increasing glacial melt and associated runoff. Although such a pattern has not been observed in other regions of the globe where runoff is predicted to decrease (Huss and Hock, 2018; St.Pierre et al., 2019b), our findings are likely to apply to other, smaller, glacierized watersheds typical of high latitudes or altitudes. Indeed, at least in the Arctic, freshwater discharge is broadly expected to increase with increasing temperatures and precipitation loadings (Peterson et al., 2002; Rawlins et al., 2010; Bring et al., 2016). It would thus be immensely valuable to conduct similar studies, replicating where appropriate a similar space-for-time design, at other lakes throughout the world. Additional sampling efforts should carefully consider the spatial heterogeneity of runoff regimes leading to divergent sedimentation rates (**Supplementary Table 3**), limiting our ability to make temporal predictions.

The hierarchical clustering showed that diversity at the phylum level generally decreased from the L to the H sites (**Figure 4**), and that the L and the C sites clustered together. This finding is unsurprising, as the L and C sites are neither distinct in terms chemical (**Figure 1B**) or taxonomic composition (**Supplementary Figures 5, 6**). Yet as the C sites are characterized by negligible runoff, this may suggest that either the L→C connection is not solely due to runoff, or that receiving low runoff temporarily increases taxonomic diversity. This observation that environments of low or intermediate disturbances can lead to maximum diversity is quite common and may provide a possible explanation for the ecological similarity between these two spatially separated sites (Gibbons et al., 2016).

Despite lacking geochemical measurements for the soil samples, we found that the microbial communities in the sediments at the high runoff sites clustered most frequently with those in the soil sites (**Figure 4**—see hierarchical clustering to the left of inset heatmaps), highlighting a potential connection between terrestrial and aquatic sediment communities as a function of the runoff volume, consistent with previous findings (Ruiz-González et al., 2015; Comte et al., 2018). In the High Arctic, glacial rivers travel across poorly consolidated landscapes with little vegetation, thus facilitating the rapid erosion and transport of materials, including soils. Glacial rivers in the High Arctic are therefore extremely turbid and discharge large loads of suspended materials and nutrients into Lake Hazen (St.Pierre et al., 2019a,b). Because of the large effect that these rivers have on the chemistry of downstream aquatic ecosystems, we would expect increased runoff to the aquatic ecosystems to alter microbial community structure (Le et al., 2016). Some of these structural changes may then alter the functional capacity to metabolize carbon, nitrogen, sulfur compounds, and process toxins such as metals and antibiotics (**Supplementary Figure 10**). A more experimentally-driven approach, based for instance on *in situ* incubation and geochemical tracers, would have been necessary to quantify such an interplay between microbial metabolism and geochemical features. Yet, as sediments and nutrients are mostly deposited during the summer melt months, it can be expected that lake sediments record microbe-driven seasonal changes in their geochemistry. Indeed, high glacial runoff is known to bring dense, oxygenated river waters with OC directly to the bottom

of the lake (St.Pierre et al., 2019b), stimulating aerobic microbial activity. As a result, the geochemistry recorded along the high runoff transect may first reflect a period of greater microbial metabolism, which may actually exceed those in temperate systems (Probst et al., 2018), eventually followed by low oxygen, low redox, and high  $\text{NH}_3$  conditions observed here (**Figure 1**) as oxygen is depleted and anaerobic metabolisms allowed to proceed.

At a larger temporal scale, a key question that arises from these results is how changes in hydrological regimes will alter the evolutionary dynamics of microbial communities in lake sediments. Niche differentiation, where the coexistence of ecological opportunities can facilitate species diversification, may explain why sediments along the low runoff transect hosts a more diverse microbial community than sediments along the high runoff transect (Cordero and Polz, 2014). Presently, climate change is predicted to increase runoff in this High Arctic environment (Lehnherr et al., 2018), and we found evidence suggesting that the increased runoff decreased diversity at the phylum level (**Figure 4**). This can be expected to disrupt niche differentiation, and hence to reduce the overall and long-term metabolic capacity in lake sediments. It is currently hard to predict the future microbial ecology of these systems. On the one hand, climate change may diminish species diversification, and lead to highly specialized microbial communities adapted to a uniform ecological niche characterized by low oxygen, low redox, and high  $\text{NH}_3$  concentrations. On the other hand, the seasonal and rapid changes in redox conditions, predicted to follow the strong but punctual input of oxygen and nutrients during springtime may allow for the development of a short-lived community that eluded our sampling and analysis.

The rapid changes that affect Lake Hazen's watershed in response to climate warming were already known to directly alter its hydrological regime. Here we further provide evidence that a combination of increasing runoff and changing geochemical conditions are associated with the reduced diversity and metabolic potential of its dominant microbial communities. While longitudinal studies are needed to confirm these patterns, it is still unclear how such losses in biodiversity and metabolic potential in Arctic ecosystems will impact key biogeochemical cycles, potentially creating feedback loops of uncertain direction and magnitude.

## MATERIALS AND METHODS

### Sample Collection and Processing

Sediment and soil cores were collected from Lake Hazen (82°N, 71°W: **Figure 1A**), located within Quttinirpaaq National Park, on northern Ellesmere Island, Nunavut. Sampling took place between May 10 and June 10, 2017, when the lake was still completely ice-covered and prior to the onset of glacial melt within the watershed (**Supplementary Table 4**). Between late June and the end of August, meltwaters flow from the outlet glaciers along the northwestern shoreline through poorly consolidated river valleys, depositing sediments at the bottom of Lake Hazen. Sediment cores were sampled along two transects, the H1/H2 and L1/L2 sites, representing areas of the lake

influenced by large (high runoff) and small (low runoff) rivers, respectively (**Supplementary Table 5**, adapted from St.Pierre et al., 2019b). The lake then drains via the Ruggles River along its southeastern shoreline (C site). The surrounding glacial rivers deliver different amounts of sediments, nutrients, and organic carbon unevenly to the lake as a consequence of heterogeneous sedimentation rates (**Supplementary Table 6**, adapted from St.Pierre et al., 2019b). More specifically, the top 5 cm of sediments from the deeper low (L2) and high (H2) runoff sites represented depositional periods of 30 years (1987–2017) and 6 years (2011–2017), respectively.

Samples were collected along two transects and can be separated into three hydrological regimes by seasonal runoff volume: low (L transect), high (H transect), and negligible runoff (C sites) summarized in **Supplementary Table 5**. Contamination of samples was minimized by wearing non-powdered latex gloves during sample handling and sterilizing all equipment with 10% bleach and 90% ethanol before sample collection. Sediment cores approximately 30 cm in length were collected with an UWITEC (Mondsee, Austria) gravity corer from five locations: C (overlying water depth: 50 m) far from the direct influence of glacial inflows serving as a control site; L1 (water depth: 50 m) and L2 (water depth: 251 m), at variable distances from a small glacial inflow (Blister Creek, <0.001 km<sup>3</sup> in summer 2016); and, H1 (water depth: 21 m) and H2 (water depth: 253 m), located adjacent to several larger glacial inflows (i.e., the Abbé River, 0.015 km<sup>3</sup> and Snow Goose, 0.006 km<sup>3</sup> in 2016). Note that bathymetry changes quickly on the northern side of Lake Hazen, making it challenging to match the shallow sites (L1 and H1) when coring through the ice-covered lake. The glacial inflow measurements were included in a previous study (St.Pierre et al., 2019b), and are summarized in **Supplementary Table 5**. The soil samples (S sites) were collected from three sites in the dried streambeds of the tributaries, on the northern shore between the two transects. At each site, for both sediments and soil, five cores were sampled, ~3 m apart for the sediment cores, and approximately ~1 m apart to account for site heterogeneity.

Microprofiling data were previously described (St.Pierre et al., 2019b). Briefly, for sediment core, one of the five cores was used for microprofiling of oxygen ( $\text{O}_2$ ), redox and pH, as well as one core for porewater chemistry and loss on ignition (see Ruuskanen et al., 2018, for details), and the remaining three cores were combined, prior to their genomic analysis, here again to account for site heterogeneity. For soil samples, three cores per site were collected for DNA analysis, but no additional cores were collected for chemical analyses. As we were mostly interested in the community composition through space, we combined the top 5 cm of sediment and 10 cm of soil for sample extraction and subsequent sequencing (below 10 cm, the ground was frozen and could not be penetrated safely by our corer). Critically, surface vegetation was scrubbed off if present and not included in the extraction vials. This was done to limit any plant DNA being captured in the extraction process. Any remaining length of cores that were used for DNA analysis were discarded. These uppermost layers in the sediment correspond to both the most recent sediment deposition dates and the region of greatest microbial activity (Haglund et al., 2003). Based on

previous findings (St.Pierre et al., 2019b), sediment below 10 cm in depth can date back as early as 2009 in the high runoff sites, or as late as 1900 in low runoff sites (**Supplementary Table 6**). Although the age of the sediment at a given sample depth can vary, we were most concerned with how the microbial community in the most active depths responded to changing environmental conditions as a result of increased runoff and sediment delivery. The top of each core was sectioned and placed into Whirlpack bags. These slices were homogenized manually inside of the bags and stored in a  $-20^{\circ}\text{C}$  freezer until shipment back to the University of Ottawa where they were then stored at  $-80^{\circ}\text{C}$ . Soil samples were transferred into falcon tubes, homogenized, and stored as described above for the lake sediment samples.

Samples were thawed overnight and 250–400 mg (wet weight; **Supplementary Table 7**) were then washed in a sterile salt buffer (10 mM EDTA, 50 mM Tris-HCl, 50 mM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$  at pH 8.0) to remove PCR inhibitors (Zhou et al., 1996; Poulain et al., 2015). All sample handling was conducted in a stainless-steel laminar flow hood (HEPA 100) treated with UVC radiation and bleach before use. DNA extractions were performed using the DNeasy PowerSoil Kit (MO BIO Laboratories Inc, Carlsbad, CA, USA), following the kit guidelines, except that the final elution volume was  $30\ \mu\text{l}$  instead of  $100\ \mu\text{l}$ . DNA integrity was validated with a NanoDrop Spectrometer and PCR combined with electrophoresis of the Glutamine synthetase gene (*glnA*) as this gene is ubiquitous across microbial life (**Supplementary Figure 15**, **Supplementary Table 8**). Adequate DNA concentrations for sequencing were reached by combining triplicate extractions for a total volume of  $45\ \mu\text{l}$  and a concentration  $\geq 50\ \text{ng}/\mu\text{l}$  (**Supplementary Table 7**). Positive and negative controls were used to verify the integrity of the PCR amplification of *glnA* (see also Ruuskanen et al., 2020). Two kit extraction blanks contained no trace of DNA and were not sequenced.

## Chemical Analyses

Redox potential, pH, and dissolved  $\text{O}_2$  concentration profiles were measured at  $100\ \mu\text{m}$  intervals on one of the cores within an hour of collection, using  $100\ \mu\text{m}$  Unisense (Aarhus, Denmark) glass microsensors connected to a Unisense Field Multimeter (OX-100, pH-100 coupled to REF-RM, RD-100 coupled to REF-RM). All sensors were calibrated immediately before the profiles were measured using the standard calibration procedures outlined in the individual sensor manuals provided by the manufacturer. A 5 mm reference electrode (Ag-AgCl; Ref-RM) was coupled with both the pH and redox potential sensors and kept in the water overlying the core during profiling. All profiles were begun at  $\sim 1\ \text{cm}$  above the sediment surface, which was approximately uniform at all sites. At each step, the probes equilibrated for 10 s before taking triplicate measurements, which were averaged to produce the profiles. A single profile was conducted on each core to measure the natural *in situ* biogeochemical gradients present at each site. We note that redox profiles should be interpreted as relative rather than

absolute differences (Boudreau and Jorgensen, 2001, p. 180–210). Sediment porewater was extracted following centrifugation at 4,000 rpm. The supernatant was then filtered through  $0.45\ \mu\text{m}$  cellulose acetate filters into 15 ml tubes, and were frozen until analysis. Concentrations of nitrates and nitrites ( $\text{NO}_2^- + \text{NO}_3^-$ ), and ammonia ( $\text{NH}_3$ ), chloride ( $\text{Cl}^-$ ) were measured in the sediment porewater using a Lachat QuickChem 8500 FIA Ion Analyzer, while total dissolved phosphorus (TDP) and  $\text{SO}_4^{2-}$  were measured in the sediment porewater using an ion chromatograph at the Biogeochemical Analytical Service Laboratory (Department of Biological Sciences, University of Alberta). However, TDP was removed from data analysis because insufficient porewater was collected to measure TDP at site C. The centrifuged sediments were retained and percentage per dry weight (% d.w.) of calcium carbonate ( $\text{CaCO}_3$ ) and organic carbon (OC) were estimated through loss on ignition (Heiri et al., 2001).

The chemical features of the top 5 cm of the sediment cores were derived from measurements performed at 1 cm intervals throughout the cores, and were reported in St.Pierre et al. (2019b) as part of a larger study on the Lake Hazen watershed, which did not investigate microbial communities in the lake. The geochemical properties of each sediment site were summarized using a Principal Component Analysis (PCA) that was scaled to unit variance and projections were clustered using Partitioning Around Medoids (Maechler et al., 2019). The appropriate number of clusters was determined from silhouettes with the R package hopach (van der Laan and Pollard, 2003). The Dunn test R package (Dinno, 2017) was used to compare samples, controlling for multiple comparisons with the Benjamini-Hochberg adjustment.

## Sequencing and Data Processing

Metagenomic libraries were prepared and sequenced by Genome Quebec on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA; **Supplementary Figure 16**) on a paired-end 125 bp configuration using Illumina TruSeq LT adapters (read 1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC, and read 2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). The DNA from the eight sites (five sediment and three soil) was sequenced with two samples per HiSeq lane, generating a minimum of 125 million reads per sample, which amounted to over 150 GB of sequencing data. Read count summaries were tracked throughout each step of the pipeline for quality control (**Supplementary Figure 17**). Low quality reads, adapters, unpaired reads, and low quality bases at the ends of reads were removed to generate quality controlled reads with Trimmomatic (v0.36) (Bolger et al., 2014) using the following arguments: `phred33, ILLUMINACLIP:TruSeq3-PE-2.fa:3:26:10, LEADING:3 TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:36, CROP:120, HEADCROP:20, AVGQUAL:20`. Then, FASTQC (v0.11.8) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to confirm that the Illumina adapters were removed and that trimmed sequence lengths were at least 90 bp in length with a Phred score of at least 33.



## Reconstruction of Environmental Genomes and Annotation

To reconstruct environmental genomes, metagenomic quality-controlled reads from all samples were coassembled using Megahit (Li et al., 2015) software with a k-mer size of 31 and meta-large setting (see **Supplementary Table 1** for additional summary statistics). EukRep (West et al., 2018) was used to remove any eukaryotic DNA from the contigs prior to the formation of an Anvio (v5) (Eren et al., 2015) contig database. The contig database was generated by removing contigs under 1000 bp, and gene prediction was performed in the Anvio environment. Sequence coverage information was determined for each assembled scaffold by mapping reads from each sample to the assembled contig database using Bowtie2 (Langmead and Salzberg, 2012) with default settings. The resulting SAM files were sorted and converted to BAM files using samtools (v0.1.19) (Li et al., 2009). Each BAM file was prepared for Anvio using the `anvi-init-bam` and contig database generated using `anvi-gen-contigs-database`. The contig database and BAM mapping files were further used as input for `anvi-profile`, which generated individual sample profiles for each contig over the minimum length of 2,500 bp. These profiles were then combined using `anvi-merge` and summary statistics for abundance and coverage were generated with `anvi-summarize`. Automated binning was performed using CONCOCT (Alneberg et al., 2014). Scaffolds were binned on the basis of GC content and differential coverage abundance patterns across all eight samples. Manual refinement was done using Anvio's `refine` option, where contigs were manually removed from bins on the basis of GC content, differential abundance in samples, and taxonomy of contigs assigned by Kaiju (Menzel et al., 2016, **Supplementary Table 2**). Kaiju was used to classify taxonomy of the assembled contigs with `anvi-import-taxonomy-for-genes` and aided in the manual refinement process. Open reading frames were predicted with Prodigal (v2.6.3) (Hyatt et al., 2010). Anvio's custom Hidden Markov Models were run, along with NCBI's COG (Tatusov et al., 2003) annotation to identify protein-coding genes. PFAM (Finn et al., 2015), TIGRFAM (Haft et al., 2003), GO terms (Ashburner et al., 2000), KEGG enzymes and pathways (Kanehisa et al., 2015), and Metacyc pathways (Caspi et al., 2007) were predicted with Interproscan (v5) (Jones et al., 2014). These annotations were then combined with the Anvio database with `anvi-import-functions`.

Genome completeness and contamination were evaluated on the presence of a core set of genes using CheckM (v1.0.5) `lineage_wf` (**Supplementary Table 2**, **Supplementary Figure 18**; Parks et al., 2015). Only the 300 genomes that satisfied the quality control cutoffs of at least 50% complete and with <10% contamination were further analyzed—meeting the Minimum Information about metagenome-assembled genome (MIMAG) of bacteria and archaea for medium or high-quality genomes (Bowers et al., 2017). All recovered genomes were used to calculate an average amino acid identity across all genomes using `compareM` (v0.0.23, function `aa_i_wf`; <https://github.com/dparks1134/CompareM>;

Parks et al., 2017). CheckM was used again to identify contigs that were not contained in any of the 300 high-quality genomes, that is those whose size ranges from 1,000 to 2,500 bp. As an attempt to “rescue” these unbinned contigs, an alternative binning algorithm MaxBin (v2.0) (Wu et al., 2015) was employed. An additional 481 genomes were recovered, but were not included in further analysis as only 21 genomes were of average completion >65% (**Supplementary Data 1**: [github.com/colbyga/hazen\\_metagenome\\_publication/blob/master/Supplemental\\_Data\\_1\\_maxbin2\\_unbinned\\_contigs\\_summary.csv](https://github.com/colbyga/hazen_metagenome_publication/blob/master/Supplemental_Data_1_maxbin2_unbinned_contigs_summary.csv)).

## Phylogenetic Placement of the MAGs

Phylogenetic analyses were performed using two different sets of marker genes from the Genome Taxonomy Database (GTDB): one for bacteria (120 marker genes) and one for archaea (122 marker genes), as previously been used to assign taxonomy to MAGs (Parks et al., 2018). The marker genes were extracted from each genome by matching Pfam (v31) (Finn et al., 2015) and TIGRFAM (v15.0) (Haft et al., 2003) annotations from GTDB (v86) (Parks et al., 2018). Marker genes from each of the 300 genomes were translated using the R package `seqinr` (Charif and Lobry, 2007), selecting the genetic code that returned no in-frame stop codon. As some genomes had multiple copies of a marker gene, duplicated copies were filtered out by keeping the most complete sequence. Marker genes that were missing from some genomes were replaced by indel (gap) characters, and their concatenated sequences were added those from the reference GTDB sequences. MUSCLE (v3.8.31) (Edgar, 2004) was employed to construct the alignment in R (v 3.5.1) (R Development Core Team, 2008). Archaeal sequences were removed from the bacterial alignment on the basis of results from CheckM (Parks et al., 2015) and independently verified using a custom list of archaea specific marker genes. Alignments were then refined using `trimAI` (Capella-Gutiérrez et al., 2009) and the `-gappyout` parameter. FastTree2 (Price et al., 2010), recompiled with double precision to resolve short branch lengths, was used to infer maximum likelihood phylogenetic trees from protein sequence alignments under the WAG + $\Gamma$  model (Whelan and Goldman, 2001; Aris-Brosou and Rodrigue, 2012, 2019). The archaeal tree was rooted with Euryarchaeota and the bacterial tree was rooted with Patescibacteria using the R package APE (Paradis et al., 2004). Trees were visualized and colored by phylum with the R package `ggtree` (Yu et al., 2017).

## Community Composition of the MAGs

To determine the relative abundance of each genome in the eight samples, sample-specific genome abundances were normalized by sequencing depth [(reads mapped to a genome)/(total number of reads mapped)], making comparisons across samples possible. Genome abundances were generated using the `CheckM profile` function (Parks et al., 2015). To determine the average abundance of major taxonomic groups across sites (determined by the phylogenetic analysis described above), the abundances for genomes from the same taxonomic group were summed and visualized using the R package `phyloseq` (McMurdie and Holmes, 2013). These same abundance values were the basis

for a community composition analysis. The *t*-SNE plots were constructed by assigning each genome to a site based on where it was most abundant using the R package Rtsne (Krijthe et al., 2018). All these analyses were made at the phylum level, unless otherwise stated.

## Metabolic Potential of the MAGs

To analyze functional marker genes in the metagenomes, we used a custom database of reference proteins sequences (COG, PFAM, TIGRFAM, KEGG) based on the marker genes used in other studies (Anantharaman et al., 2016; Dombrowski et al., 2018, **Supplementary Data Files** on GitHub). Pathways were also predicted using MinPath (Ye and Doak, 2009) to map all identified KEGG enzymes to the most parsimonious MetaCyc pathways (Caspi et al., 2007). As these MAGs were incomplete, some genes in pathways may be absent. MinPath presented only parsimonious pathways represented by multiple genes. As most genomes were present even at low abundances across all sites, a cut-off value of  $\leq 0.25$  (on a  $-\log_{10}$  scale) was set for a genome to be included in the functional analyses at any site, so that only the most abundant genomes for each site were considered. We aggregated marker genes and pathways by function, summarizing the results by phyla, except for Proteobacteria that were separated by class. We further grouped all taxa together at each site to test for significant differences in major nutrient cycling processes (carbon, nitrogen, and sulfur) among sites using a hierarchical clustering; significance was derived from the Approximately Unbiased bootstrap (Suzuki and Shimodaira, 2006) and Fisher's exact test.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556841>.

## REFERENCES

- Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11:1144. doi: 10.1038/nmeth.3103
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- Aris-Brosou, S., and Rodrigue, N. (2012). "The essentials of computational molecular evolution," in *Evolutionary Genomics* (Totowa, NJ: Humana Press), 111–152.
- Aris-Brosou, S., and Rodrigue, N. (2019). A not-so-long introduction to computational molecular evolution. *Methods Mol. Biol.* 1910, 71–117. doi: 10.1007/978-1-4939-9074-0\_3
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25. doi: 10.1038/75556
- Bliss, A., Hock, R., and Radić, V. (2014). Global response of glacier runoff to twenty-first century climate change. *J. Geophys. Res.* 119, 717–730. doi: 10.1002/2013JF002931

## AUTHOR CONTRIBUTIONS

GC and VS performed sampling, whereas GC and MR conducted laboratory analyses. GC, MR, and SA-B performed data analyses. GC, SA-B, VS, and AP designed the study and wrote the manuscript. KS and VS conducted the microsensor profiles and porewater extractions. GC, SA-B, AP, MR, KS, and VS reviewed the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

Funding support was provided by the Natural Science and Engineering Research Council of Canada (VS, AP, and SA-B), ArcticNet Network Center of Excellence (VS and AP), and the Polar Continental Shelf Program (VS) in Resolute, Nunavut, which provided logistical and financial support.

## ACKNOWLEDGMENTS

This study was made possible through a collaborative effort undertaken by Igor Lehnerr, Stephanie Varty, Victoria Wisniewski (University of Toronto, Mississauga), Charles Talbot (Environment and Climate Change Canada), and Maria Cavaco (University of Alberta). We thank Linda Bonen, Marina Cvetkovska, Manon Ragonnet, and Alex Wong for comments and discussions, and three reviewers for highly constructive comments. This manuscript has been released as a pre-print at bioRxiv (Colby et al., 2020).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.561194/full#supplementary-material>

- Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., and Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9374–9379. doi: 10.1073/pnas.1220228110
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boudreau, B. P., and Jorgensen, B. B. (2001). *The Benthic Boundary Layer: Transport Processes and Biogeochemistry*. Oxford: Oxford University Press.
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35:725. doi: 10.1038/nbt.3893
- Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S., et al. (2016). Arctic terrestrial hydrology: a synthesis of processes, regional effects, and research challenges. *J. Geophys. Res.* 121, 621–649. doi: 10.1002/2015JG003131
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., et al. (2007). The MetaCyc database of metabolic pathways and enzymes

- and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 36, D623–D631. doi: 10.1093/nar/gkm900
- Cavaco, M. A., St. Louis, V. L., Engel, K., St. Pierre, K. A., Schiff, S. L., Stibal, M., et al. (2019). Freshwater microbial community diversity in a rapidly changing high arctic watershed. *FEMS Microbiol. Ecol.* 95. doi: 10.1093/femsec/fiz161
- Charif, D., and Lobry, J. (2007). “SeqinR 1.0.2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis,” in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, eds U. Bastolla, M. Porto, H. Roman, and M. Vendruscolo (New York, NY: Springer Verlag), 207–232.
- Colby, G. A., Ruuskanen, M. O., St. Pierre, K. A., St. Louis, V. L., Poulain, A. J., Aris-Brosou, S. (2020). Climate change negatively impacts dominant microbes in the sediments of a High Arctic lake. *bioRxiv* doi: 10.1101/705178
- Comte, J., Culley, A. L., Lovejoy, C., and Vincent, W. F. (2018). Microbial connectivity and sorting in a High Arctic watershed. *ISME J.* 12, 2988–3000. doi: 10.1038/s41396-018-0236-4
- Cordero, O. X., and Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12:263. doi: 10.1038/nrmicro3218
- Crevecoeur, S., Vincent, W. F., Comte, J., and Lovejoy, C. (2015). Bacterial community structure across environmental gradients in permafrost thaw ponds: methanotroph-rich ecosystems. *Front. Microbiol.* 6:192. doi: 10.3389/fmicb.2015.00192
- Crump, B. C., Amaral-Zettler, L. A., and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J.* 6, 1629–1639. doi: 10.1038/ismej.2012.9
- Damgaard, C. (2019). A critique of the space-for-time substitution practice in community ecology. *Trends Ecol. Evol.* 34, 416–421. doi: 10.1016/j.tree.2019.01.013
- Dinno, A. (2017). *Dunn's Test of Multiple Comparisons Using Rank Sums*. R package version 1.
- Dombrowski, N., Teske, A. P., and Baker, B. J. (2018). Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* 9:4999. doi: 10.1038/s41467-018-07418-0
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvio: an advanced analysis and visualization platform for omics data. *PeerJ* 3:1319. doi: 10.7717/peerj.1319
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Frey, K. E., and McClelland, J. W. (2009). Impacts of permafrost degradation on Arctic river biogeochemistry. *Hydrol. Process.* 23, 169–182. doi: 10.1002/hyp.7196
- Gibbons, S. M., Scholz, M., Hutchison, A. L., Dinner, A. R., Gilbert, J. A., and Coleman, M. L. (2016). Disturbance regimes predictably alter diversity in an ecologically complex bacterial system. *Mbio* 7, e01372–e01316. doi: 10.1128/mBio.01372-16
- Haft, D. H., Selengut, J. D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. doi: 10.1093/nar/gkg128
- Haglund, A.-L., Lantz, P., Törnblom, E., and Tranvik, L. (2003). Depth distribution of active bacteria and bacterial activity in lake sediment. *FEMS Microbiol. Ecol.* 46, 31–38. doi: 10.1016/S0168-6496(03)00190-9
- Heiri, O., Lotter, A. F., and Lemcke, G. (2001). Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *J. Paleolimnol.* 25, 101–110. doi: 10.1023/A:1008119611481
- Hermans, S. M., Buckley, H. L., Case, B. S., and Lear, G. (2020). Connecting through space and time: catchment-scale distributions of bacteria in soil, stream water and sediment. *Environ. Microbiol.* 22, 1000–1010. doi: 10.1111/1462-2920.14792
- Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., et al. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521:208. doi: 10.1038/nature14238
- Huss, M., and Hock, R. (2018). Global-scale hydrological response to future glacier mass loss. *Nat. Clim. Change* 8:135. doi: 10.1038/s41558-017-0049-x
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Köck, G., Muir, D., Yang, F., Wang, X., Talbot, C., Gantner, N., et al. (2012). Bathymetry and sediment geochemistry of Lake Hazen (Quttinirpaaq National Park, Ellesmere Island, Nunavut). *Arctic* 56–66. doi: 10.14430/arctic4165
- Krijthe, J., van der Maaten, L., and Krijthe, M. J. (2018). *Package 'Rtsne'*.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923
- Laudon, H., Spence, C., Buttle, J., Carey, S. K., McDonnell, J. J., McNamara, J. P., et al. (2017). Save northern high-latitude catchments. *Nat. Geosci.* 10, 324–325. doi: 10.1038/ngeo2947
- Le, H. T., Ho, C. T., Trinh, Q. H., Luu, M. T., Tran, H. S., et al. (2016). Responses of aquatic bacteria to terrestrial runoff: effects on community structure and key taxonomic groups. *Front. Microbiol.* 7:889. doi: 10.3389/fmicb.2016.00889
- Lehnherr, I., Louis, V. L. S., Sharp, M., Gardner, A. S., Smol, J. P., Schiff, S. L., et al. (2018). The world's largest High Arctic lake responds rapidly to climate warming. *Nat. Commun.* 9:1290. doi: 10.1038/s41467-018-03685-z
- Lester, R. E., Close, P. G., Barton, J. L., Pope, A. J., and Brown, S. C. (2014). Predicting the likely response of data-poor ecosystems to climate change using space-for-time substitution across domains. *Glob. Change Biol.* 20, 3471–3481. doi: 10.1111/gcb.12634
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lynch, M. D. J., and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229. doi: 10.1038/nrmicro3400
- Mackelprang, R., Saleska, S. R., Jacobsen, C. S., Jansson, J. K., and Taş, N. (2016). Permafrost meta-omics and climate change. *Annu. Rev. Earth Planet. Sci.* 44, 439–462. doi: 10.1146/annurev-earth-060614-105126
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). *cluster: Cluster Analysis Basics and Extensions*. R package version 1, 56.
- McCalley, C. K., Woodcroft, B. J., Hodgkins, S. B., Wehr, R. A., Kim, E.-H., Mondav, R., et al. (2014). Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* 514:478. doi: 10.1038/nature13798
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. doi: 10.1371/journal.pone.0061217
- Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:11257. doi: 10.1038/ncomms11257
- Mingle, J. (2020). *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. New York, NY.
- Mohit, V., Culley, A., Lovejoy, C., Bouchard, F., and Vincent, W. F. (2017). Hidden biofilms in a far northern lake and implications for the changing Arctic. *NPJ Biofilms Microbiomes* 3:17. doi: 10.1038/s41522-017-0024-3
- O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., et al. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophys. Res. Lett.* 42, 10–773. doi: 10.1002/2015GL066235
- Overpeck, J., Hughen, K., Hardy, D., Bradley, R., Case, R., Douglas, M., et al. (1997). Arctic environmental change of the last four centuries. *Science* 278, 1251–1256. doi: 10.1126/science.278.5341.1251
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome

- phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2:1533. doi: 10.1038/s41564-017-0012-7
- Peterson, B. J., Holmes, R. M., McClelland, J. W., Vörösmarty, C. J., Lammers, R. B., Shiklomanov, A. I., et al. (2002). Increasing river discharge to the arctic ocean. *Science* 298, 2171–2173. doi: 10.1126/science.1077445
- Pickett, S. T. (1989). “Space-for-time substitution as an alternative to long-term studies,” in *Long-Term Studies in Ecology* (New York, NY: Springer), 110–135.
- Poulain, A. J., Aris-Brosou, S., Blais, J. M., Brazeau, M., Keller, W. B., and Paterson, A. M. (2015). Microbial DNA records historical delivery of anthropogenic mercury. *ISME J.* 9:2541. doi: 10.1038/ismej.2015.86
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490
- Probst, A. J., Ladd, B., Jarett, J. K., Geller-McGrath, D. E., Sieber, C. M., Emerson, J. B., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3:328. doi: 10.1038/s41564-017-0098-y
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rawlins, M. A., Steele, M., Holland, M. M., Adam, J. C., Cherry, J. E., Francis, J. A., et al. (2010). Analysis of the arctic system for freshwater cycle intensification: observations and expectations. *J. Clim.* 23, 5715–5737. doi: 10.1175/2010JCLI3421.1
- Ruiz-González, C., Niño-García, J. P., and del Giorgio, P. A. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecol. Lett.* 18, 1198–1206. doi: 10.1111/ele.12499
- Ruuskanen, M. O., Aris-Brosou, S., and Poulain, A. J. (2020). Swift evolutionary response of microbes to a rise in anthropogenic mercury in the Northern hemisphere. *ISME J.* 14, 788–800. doi: 10.1038/s41396-019-0563-0
- Ruuskanen, M. O., Colby, G., St. Pierre, K. A., St. Louis, V. L., Aris-Brosou, S., and Poulain, A. J. (2019). Microbial genomes retrieved from high arctic lake sediments encode for adaptation to cold and oligotrophic environments. *Limnol. Oceanogr.* 65, S233–S247. doi: 10.1101/724781
- Ruuskanen, M. O., St. Pierre, K. A., St. Louis, V. L., Aris-Brosou, S., and Poulain, A. J. (2018). Physicochemical drivers of microbial community structure in sediments of Lake Hazen, Nunavut, Canada. *Front. Microbiol.* 9:1138. doi: 10.3389/fmicb.2018.01138
- Screen, J. A., and Simmonds, I. (2010). The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* 464:1334. doi: 10.1038/nature09051
- Serreze, M. C., and Francis, J. A. (2006). The Arctic amplification debate. *Clim. change* 76, 241–264. doi: 10.1007/s10584-005-9017-y
- Smol, J. P., Wolfe, A. P., Birks, H. J. B., Douglas, M. S., Jones, V. J., Korhola, A., et al. (2005). Climate-driven regime shifts in the biological communities of Arctic lakes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4397–4402. doi: 10.1073/pnas.0500245102
- St. Pierre, K., St. Louis, V., Lehnerr, I., Schiff, S., Muir, D., Poulain, A., et al. (2019b). Contemporary limnology of the rapidly changing glacierized watershed of the world’s largest high arctic lake. *Sci. Rep.* 9:4447. doi: 10.1038/s41598-019-39918-4
- St. Pierre, K. A., St. Louis, V. L., Schiff, S. L., Lehnerr, I., Dainard, P. G., Gardner, A. S., et al. (2019a). Proglacial freshwaters are significant and previously unrecognized sinks of atmospheric CO<sub>2</sub>. *Proc. Natl. Acad. Sci. U.S.A.* 116, 17690–17695. doi: 10.1073/pnas.1904241116
- Stoeva, M. K., Aris-Brosou, S., Chételat, J., Hintelmann, H., Pelletier, P., and Poulain, A. J. (2014). Microbial community structure in lake and wetland sediments from a High Arctic polar desert revealed by targeted transcriptomics. *PLoS ONE* 9:e89531. doi: 10.1371/journal.pone.0089531
- Suzuki, R., and Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Thaler, M., Vincent, W. F., Lionard, M., Hamilton, A. K., and Lovejoy, C. (2017). Microbial community structure and interannual change in the last epishelf lake ecosystem in the north polar region. *Front. Mar. Sci.* 3:275. doi: 10.3389/fmars.2016.00275
- van der Laan, M. J., and Pollard, K. S. (2003). Hybrid clustering of gene expression data with visualization and the bootstrap. *J. Stat. Plann. Inference.* 117, 275–303. doi: 10.1016/S0378-3758(02)00388-9
- West, P. T., Probst, A. J., Grigoriev, I. V., Thomas, B. C., and Banfield, J. F. (2018). Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.* 28, 569–580. doi: 10.1101/gr.228429.117
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Williamson, C. E., Saros, J. E., Vincent, W. F., and Smol, J. P. (2009). Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnol. Oceanogr.* 54, 2273–2282. doi: 10.4319/lo.2009.54.6\_part\_2.2273
- Wrona, F. J., Johansson, M., Culp, J. M., Jenkins, A., Mård, J., Myers-Smith, I. H., et al. (2016). Transitions in Arctic ecosystems: ecological implications of a changing hydrological regime. *J. Geophys. Res.* 121, 650–674. doi: 10.1002/2015JG003133
- Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638
- Ye, Y., and Doak, T. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5:e1000465. doi: 10.1371/journal.pcbi.1000465
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322. doi: 10.1128/AEM.62.2.316-322.1996

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Colby, Ruuskanen, St. Pierre, St. Louis, Poulain and Aris-Brosou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.