scientific reports

OPEN

MinION sequencing from sea ice cryoconites leads to de novo genome reconstruction from metagenomes

Catherine Maggiori✉, Isabelle Raymond-Bouchard, Laura Brennan, David Touchette & Lyle Whyte

Genome reconstruction from metagenomes enables detailed study of individual community members, their metabolisms, and their survival strategies. Obtaining high quality metagenome-assembled genomes (MAGs) is particularly valuable in extreme environments like sea ice cryoconites, where the native consortia are recalcitrant to culture and strong astrobiology analogues. We evaluated three separate approaches for MAG generation from Allen Bay, Nunavut sea ice cryoconites—HiSeq-only, MinION-only, and hybrid (HiSeq + MinION)—where field MinION sequencing yielded a reliable metagenome. The hybrid assembly produced longer contigs, more coding sequences, and more total MAGs, revealing a microbial community dominated by Bacteroidetes. The hybrid MAGs also had the highest completeness, lowest contamination, and highest N50. A putatively novel species of *Octadecabacter* is among the hybrid MAGs produced, containing the genus's only known instances of genomic potential for nitrate reduction, denitrification, sulfate reduction, and fermentation. This study shows that the inclusion of MinION reads in traditional short read datasets leads to higher quality metagenomes and MAGs for more accurate descriptions of novel microorganisms in this extreme, transient habitat and has produced the first hybrid MAGs from an extreme environment.

Cryoconites are small holes (< 1 m in diameter, < 0.5 m deep) on icy surfaces containing water and wind-blown particles[1,2]. They form when organic and inorganic materials are deposited and melt small pockets on the ice, forming a basal level of dark sediment overlain with meltwater and providing a refuge for microorganisms. Cryoconites occur in Arctic, Antarctic, and alpine regions on glaciers, ice sheets, sea ice, and lake ice[1,3–5]. They are hotspots of microbial diversity in icy environments, containing a wide variety of taxa (e.g. Cyanobacteria, heterotrophic bacteria, protists, algae, micro-invertebrates)[6] and provide key functions in icy ecosystems (e.g. carbon fixation, mineral aggregation, nutrient cycling, pollutant degradation)[1,7]. However, regional and global interaction scales of cryoconites are not well-understood[2], particularly for cryoconites on sea ice and in the Canadian Arctic Archipelago. Characterizing the microbial community in cryoconite holes will increase our comprehension of life in cold environments, for which a unified picture is severely lacking[8], and offer predictions on how changing sea ice will affect these communities, their distribution, and the surrounding ecosystem. For example, as the scope of brine in the central Arctic Ocean (CAO) expands with climate change, it is expected that freshwater and brackish Actinobacteria, Betaproteobacteria and Flavobacteriia will increase in number and range across the CAO[9]. Consequences of this expansion could be increased photoheterotrophy, increased use of dissolved organic matter (DOM), and increased infection numbers of ice-associated fish by *Flavobacterium* and *Polaribacter* in the water column[9].

Field investigations of remote locations like sea ice benefit our understanding of the microbial interactions in these fragile and increasingly transient environments[8]. Field sequencing reduces logistical risks including contamination or sample loss during transport, optimizes the capture of specific communities in space and time[10], and can be performed with the Oxford Nanopore Technologies' (ONT) MinION sequencer, a miniaturized device designed to sequence DNA, RNA, and potentially proteins[11–13]. The MinION also generates sequences in real-time, enabling rapid identification of community members for further targeted analyses in the field (e.g. determining if metabolic genes present are active via in situ respiration). It is low cost, has low energy

Department of Natural Resource Sciences, Faculty of Agricultural and Environmental Sciences, McGill University, 21 111 Lakeshore Road, Macdonald Stewart Building, Room MS3-053, Ste. Anne-de-Bellevue, Quebec H9X 3V9, Canada. ✉email: catherine.maggiori@gmail.com

requirements, and can sequence samples containing environmental inhibitors and low biomass[12,14]. The MinION has been successfully used in extreme environments, generating sequences from Svalbard, Axel Heiberg Island, Antarctica, and the International Space Station[12,15–17], and the MinION is being explored as a tool for direct life detection in astrobiology studies. DNA is an unambiguous biosignature and the MinION has potential to be adapted and utilized for in situ life detection[12,18–20]. Testing the MinION in a variety of Mars analogue environments further establishes its utility in this context. Cryoconites are analogues to icy extraterrestrial environments (e.g. Martian polar ice caps, Enceladus, Europa) given their extreme environmental conditions (e.g. freezing temperatures, low nutrient input, high solar radiation)[21,22]. In particular, the icy moon Europa is a promising target in the search for biosignatures based on its possession of chaos regions: geologically young surface regions where active resurfacing may be ongoing, potentially depositing endogenous material onto the icy surface in a manner similar to cryoconite formation, where it remains exposed and accessible for future mission investigations (e.g. Europa Lander concept mission)[23,24]. The microorganisms inhabiting cryoconite holes can therefore act as promising models for astrobiology[21,22] and detailed studies of their genomes beyond initial biosignature detection with the MinION can yield valuable insights into survival strategies, lifestyles, and potential biosignatures in icy extraterrestrial analogues.

A strong benefit of single-molecule sequencers like the MinION is their ability to produce genome-sized contigs and long reads that span repetitive regions[25,26]. Assembly, genome reconstruction, and characterizing structural variations become easier with longer, more contiguous reads, but nanopore sequencing exhibits significantly higher error rates than second-generation sequencing platforms (e.g. ~ 7.5–14.5% for MinION sequencing[27], < 0.1% for Illumina sequencing[28]), although these error rates are continually being improved. Short Illumina reads have the opposite problem: highly accurate sequencing in non-repetitive regions, but low contiguity across assemblies[26]. The two types of sequencing can be combined and assembled together to produce a hybrid dataset; long MinION reads provide contiguity and repeat resolution, while short reads facilitate local base pair accuracy[29]. Hybrid assemblies have produced more accurate and more contiguous genomes than either constituent dataset alone, including for bacterial isolates (e.g. plastic and repetitive genomes of *Enterobacteriaceae*[29], plasmid-rich *Klebsiella pneumoniae*[30], GC-variable strains[25]) and eukaryotic organisms (e.g. wild tomato *Solanum pennellii*[31], winged Antarctic midge *Parochlus steinenii*[32], clownfish *Amphiprion ocellaris*[33], blacklip abalone *Haliotis rubra*[34]).

Reconstructing genomes from metagenomes—metagenome-assembled genomes (MAGs) or genome bins—is more complex than single genome reconstruction due to the greater diversity of genomes present and the introduction of intergenomic repeats i.e. similar repetitive regions from different genomes[35]. Hybrid assemblies produce metagenome datasets with more accurate community representation and more contiguous MAGs in both mock bacterial communities[36,37] and natural consortia (e.g. gut microbiomes[38], aquifers[39], wastewater[40]). However, this approach has yet to be tested in a low biomass, extreme, and remote sample site, where the reconstruction of individual community members would provide a robust understanding of their survival strategies and contributions to the local ecosystem[39].

In this study, we collected material from cryoconite holes on sea ice in Allen Bay, Nunavut in the Canadian high Arctic (Fig. 1) and performed field sequencing with the MinION. DNA extractions and sequencing runs were performed in the field in order to test the performance and robustness of portable extraction devices in tandem with the MinION in an extreme sea ice environment[41]. The MinION reads were later assembled with Illumina (HiSeq) reads to produce a hybrid metagenome and hybrid MAGs. We compared these hybrid datasets to HiSeq-only and MinION-only metagenomes and MAGs to evaluate their contiguity and accuracy (e.g. MAG completeness, contamination, N50). This comparison will determine if hybrid assembly is a viable approach for studying low biomass, extreme, and progressively more transient environments and produce the first hybrid MAGs from an extreme environment.

Results

Geochemistry of the Allen Bay sea ice collection site.
Physico-chemical parameters of the Allen Bay sea ice collection site were measured in situ in July 2018 (Table 1). The pH of the cryoconite meltwater was recorded as 7.96, with a salinity of 0.13 ppt. Total dissolved solids and dissolved oxygen were recorded as 0.173 g/L and 12.90 mg/L (91.7%), respectively. Oxidation reduction potential (ORP) was − 114.4 mV, indicating a reducing environment in the meltwater[42]. Total organic carbon (TOC) was determined to be 112.75 ppm and ammoniacal nitrogen was measured at 1.81 ppm in neighbouring cryoconite meltwater.

The Allen Bay sea ice cryoconite metagenome is dominated by bacteroidetes.
Three assembly methods were tested in this study and used to evaluate the cryoconite metagenomes and resulting metagenome-assembled genomes (MAGs): HiSeq sequencing with metaSPAdes assembly, hybrid assembly with metaSPAdes and combined HiSeq + MinION datasets, and MinION sequencing with Canu assembly (Table 2). HiSeq sequencing produced 22.9 Gbp of data, ~ 19× more than that with MinION sequencing (1.2 Gbp). Despite having a ~ 1/19 fraction of the raw data produced by the HiSeq-only assembly, the MinION-only dataset had the longest average contig length and the highest proportion of ultra-long (> 50,000 bp) contigs in its dataset. The addition of MinION data in the hybrid dataset likewise increased the average contig length, number of ultra-long contigs (> 50,000 bp), number of coding sequences classified by JGI IMG/M ER, and total MAGs produced with respect to the HiSeq-only assembly. These improvements are detailed in Table 3 and Supplementary Fig. 1.

The 3 metagenomes (HiSeq, hybrid, and MinION) differ in proportion and breakdown of taxa due to the differences in their constituent sequencing technologies[26–28]. MinION field sequencing produced a metagenome chiefly dominated (> 90%) by Bacteroidetes as Flavobacteriales (*Flavobacterium* sp. ALD4, *Flavobacterium* sp. ACAM 123, *Flavobacterium frigoris*, and *Flavobacterium gillisiae*) (Fig. 2). Like many *Flavobacterium* spp., *F.* sp.
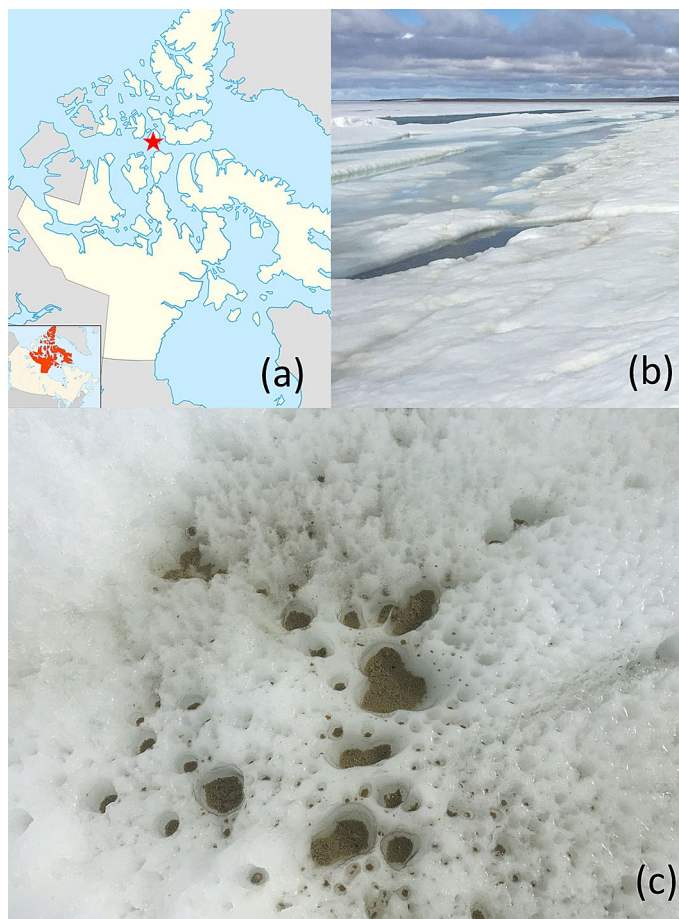
Figure 1. Allen Bay sea ice location and cryoconites. (**a**) Sampling location in the Canadian high Arctic (latitude: 74.44707; longitude: −95.0348). Map generated by Google maps version 10.86.1 (https://www.google.com/maps/place/Resolute,+NU/@76.3182414,-114.8698166,4z/data=!4m5!3m4!1s0x51dc1af3fbf4579b:0x986b529216b123b6!8m2!3d74.697299!4d-94.8297289). (**b**) Ground view of the Allen Bay sea ice. Photo taken on site by David Touchette. (**c**) Aerial view of the cryoconite sampling location. Photo taken on site by Isabelle Raymond-Bouchard.

| pH	Temperature (°C)	Salinity (ppt)	Total dissolved solids (g/L)	Dissolved oxygen (%; mg/L)	Oxidation reduction potential (ORP) (mV)	Total organic carbon (TOC) (ppm)	NH_4-N (ppm)
7.96	1.33	0.13	0.173	91.7; 12.90	−114.4	112.75	1.81

Table 1. Physico-chemical characteristics of the Allen Bay sea ice cryoconite collection site.

Assembly type	Sequencing data (Gbp)	Number of contigs	Average contig length (bp)	Number of contigs > 50,000 bp	Number of coding sequences classified by JGI	Total number of genes present	MAGs produced
HiSeq	22.9	2,197,534	1067	20	378,987	382,480	37
Hybrid	N/A	2,159,929	1163	48	397,048	400,814	44
MinION	1.2	3741	6734	36	48,475	49,539	4

Table 2. Sequencing information for each assembly method (HiSeq with metaSPAdes, hybrid with metaSPAdes and combined HiSeq/MinION datasets, MinION with Canu).

Assembly type	MAG ID	Genome size (bp)	Longest contig (bp)	Mean contig length (bp)	N50	Completeness (%)	Contamination (%)	Closest Metaerg GTDB-based taxonomy
HiSeq	HiSeq_31	3,218,255	51,014	7926	9977	90.6	1.7	*Octadecabacter arcticus*
	HiSeq_14	3,193,410	34, 424	5592	7860	80.2	3.4	*Polaromonas* sp. JS666
	HiSeq_32	2,854,964	29,037	6781	3796	77.3	0.6	*Pseudomonas* sp.
	HiSeq_30	1,577,574	14,969	4612	5723	52.1	1.7	*Nonlabens marinus*
	HiSeq_8	1,930,594	28,420	5274	4775	51.7	3.4	*Flavobacterium* sp.
	HiSeq_19	1,677,201	32,650	5392	5599	51.3	1.7	*Nonlabens* sp.
Hybrid	Hybrid_5	3,275,525	122,374	14,557	19,997	90.9	0.6	*Octadecabacter arcticus*
	Hybrid_20	3,024,995	47,365	9166	11,848	84.2	1.0	*Pseudomonas* sp.
	Hybrid_35	3,189,002	54,478	8032	10,334	79.5	1.4	*Polaromonas* sp. JS666
	Hybrid_12	1,885,346	36,215	7141	11,449	64.0	1.6	*Nonlabens marinus*
	Hybrid_27	1,821,856	32,650	6093	19,324	56.3	0.6	*Nonlabens* sp.
	Hybrid_21	2,613,398	43,486	8799	6309	50.8	0.8	*Flavobacterium* sp.

Table 3. MAGs produced with HiSeq and hybrid assembly with > 50% completeness and < 10% contamination. MAG parameters were determined with CheckM. Metaerg taxonomy is determined via predicted ORFs that inherited their taxonomy from GTDB. Taxonomy is determined to the species level if coding sequences align to one species > 50%.
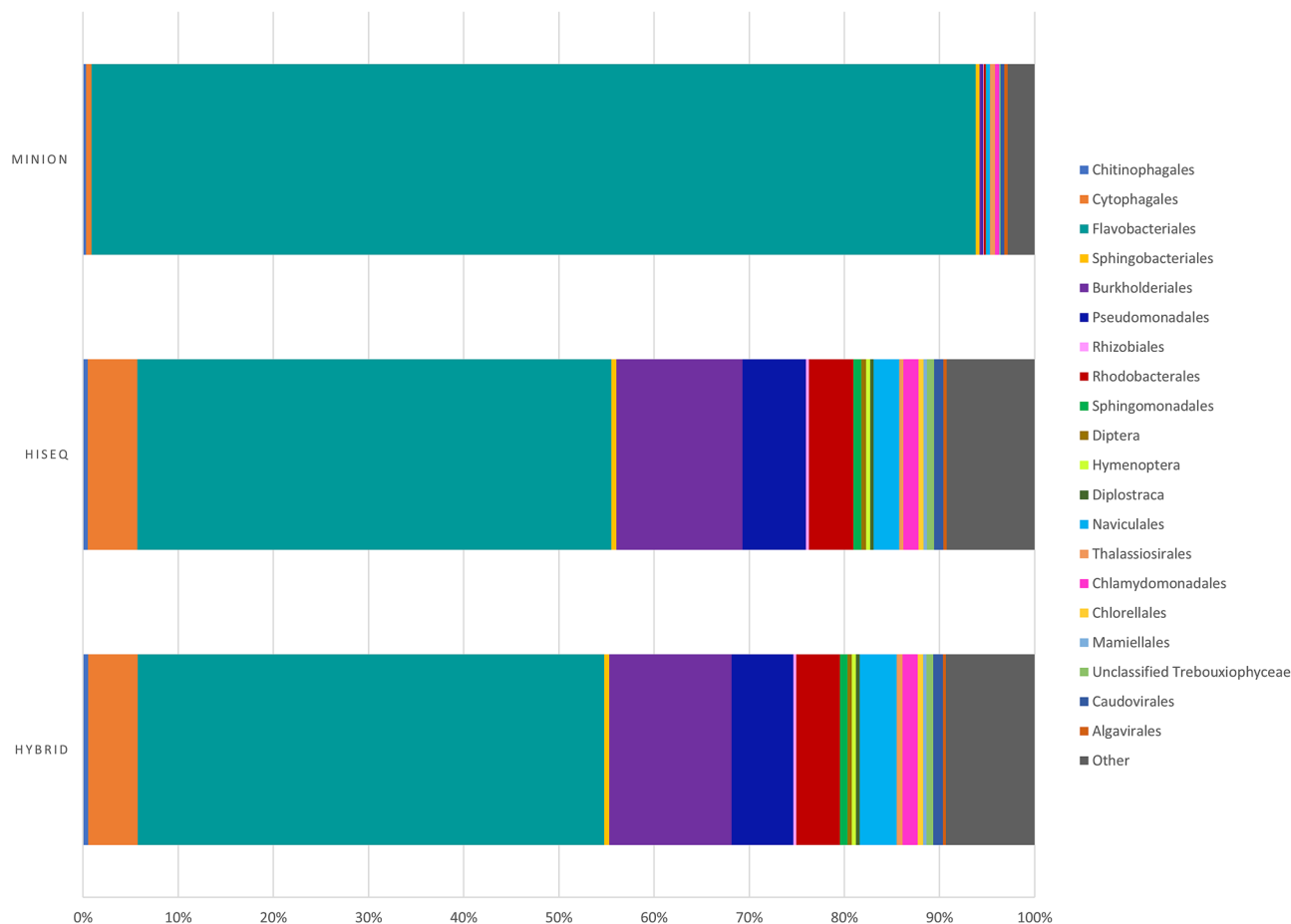
Figure 2. Community composition from metagenomes of the Allen Bay sea ice. The HiSeq-only metagenome, hybrid metagenome, and MinION-only metagenome are presented.

ALD4., *F.* sp. ACAM 123, *F. frigoris*, and *F. gillisiae* are aquatic bacterial heterotrophs and are common in polar environments (e.g. Arctic sea ice, Antarctic saline lake water, and Antarctic microbial mats)[43–45]. The metabolic pathways present in the MinION metagenome indicate largely aerobic metabolisms via oxidative phosphorylation, glycolysis, the tricarboxylic acid (TCA) cycle, the Enter-Doudoroff pathway, and glyoxylate shunt. Some fermentative and anaerobic terminal electron acceptor pathways are also present (e.g. full pathways for dissimilatory nitrate reduction, assimilatory nitrate reduction, and assimilatory sulfate reduction), as well as partial carbon fixation pathways via the reductive TCA cycle.

Similarly, the HiSeq and hybrid metagenomes largely contain Bacteroidetes (~ 55% in the HiSeq and hybrid datasets, > 90% in the MinION dataset) as Flavobacteriales (*Flavobacterium* spp.) (Fig. 2) and aerobic heterotrophic metabolic pathways (e.g. oxidative phosphorylation, the TCA cycle, the Enter-Doudoroff pathway, glyoxylate shunt). Secondarily present in the HiSeq and hybrid metagenomes are Proteobacteria as Burkholderiales (*Polaromonas* spp.), Pseudomonadales (*Pseudomonas* spp. and *Psychrobacter* spp.), and Rhodobacterales (*Loktanella salsilacus* and *Octadecabacter* spp.). Eukarya are also present as green algae (Chlorophyta), diatoms (Bacillariophyta), and arthropods (Diptera, Diplostraca), as well as small amounts of viruses (Caudovirales, Algavirales) and Archaea in all three assembly methods. Caudovirales are bacteriophages while Algavirales prey on eukaryotic algae[46]. Archaeal sequences are primarily Euryarchaeota as Methanomicrobia and Halobacteria. The HiSeq and hybrid metagenomes differ slightly in the proportions of taxa present (e.g. Bacteroidetes present as ~ 55.9% and 56.7% in the hybrid and HiSeq metagenomes, respectively; Proteobacteria present as ~ 26% and ~ 27% in the hybrid and HiSeq metagenomes, respectively), but possess all of the same taxa at the phylum level, except for Candidatus Gracilibacteria, detected exclusively in the hybrid and MinION metagenomes. Gracilibacteria are an uncultured lineage previously detected in deep-sea sediment and microbial mats[47] with limited metabolisms and an opal stop codon encoding for glycine[48].

Metagenome-assembled genome (MAG) properties from HiSeq, hybrid, and MinION assemblies. 37 and 44 MAGs were produced from the HiSeq and hybrid assemblies, respectively. Neither method produced bins with unique taxonomic assignments. Table 3 presents the details of bins with at least 50% completeness and less than 10% contamination as determined by CheckM. These details are plotted in Supplementary Fig. 1. One high-quality (i.e. > 90% complete, < 5% contaminated, presence of the 23S, 16S, and 5S rRNA genes and at least 18 tRNAs)[49] and five medium-quality bins (i.e. ≥ 50% complete, < 5% contaminated)[39,49] were produced from the hybrid method. Six medium-quality bins were obtained from the HiSeq method. Similar taxonomies for the high- and medium-quality bins were produced with each method as identified with GTDB (One *Octadecabacter* MAG, one *Polaromoas* MAG, one *Pseudomonas* MAG, two *Nonlabens* MAGs, and one *Flavobacterium* MAG), albeit with differing completeness, contamination, and genome dimensions. When directly comparing MAGs with the same taxonomy, in nearly all cases the hybrid MAGs had higher completeness, N50 values, mean contig length, longest contig, and lower contamination; the only exception is Hybrid_35 with a completeness of 79.5%, slightly lower than HiSeq_14's completeness of 80.2%. The hybrid MAGs also have larger genomes, likewise with the exception of Hybrid_35 and HiSeq_14. The taxonomies produced strongly reflect the dominance of Flavobacteriales (*Flavobacterium* sp. and *Nonlabens* spp.), Burkholderiales (*Polaromonas* sp000751355), Pseudomonadales (*Pseudomonas* sp.), and Rhodobacterales (*Octadecabacter arcticus*) in the metagenomes. Full sets of rRNA (i.e. 5S, 23S, and 16S) were present in 4 MAGs: one hybrid MAG (Hybrid_5) and three MinION MAGs (MinION_3, MinION_RD_2, MinION_RD_3).

Genome binning was attempted for the MinION-only datasets; however, the resulting MAGs were of high contamination and medium-to-low completeness, even after short-read polishing and frameshift correction (Supplementary Table 1). Only one MinION MAG could be assigned taxa below Bacteria (*Flavobacterium* sp.); however, this MAG's contamination was so high (93.9%) as to make this designation irrelevant. Nevertheless, the MinION metagenome clearly represented the taxa present in the HiSeq and hybrid MAGs (*Octadecabacter*, *Polaromonas*, *Pseudomonas*, and *Flavobacterium*).

As MinION reads are known to be of poorer quality than Illumina-produced sequences (e.g. more insertion-deletions)[27], we assessed if the addition of MinION sequences significantly affected the contig and individual MAG quality of the hybrid dataset as compared with the HiSeq dataset via the number of indels present. Higher indel rates can introduce premature stop codons and result in truncated ORFs, which can be seen in gene prediction tools: the ratio between the length of predicted proteins and their best matches will be < 1 if there are many indels present in the dataset[50]. Plots of frequency vs. query length:hit length for each assembly method, as well the highest quality HiSeq and hybrid MAGs, are presented in Fig. 3. The hybrid plots do not differ significantly from the HiSeq plots and both show that the majority of the contigs present have a query length:hit length ratio of ~ 1; the query length (our sequence) and the length of its best match (hit length) are generally the same, indicating that the effect of indels in the hybrid and HiSeq datasets is minimal. This trend is further reflected in the highest quality hybrid and HiSeq MAGs (Hybrid_5 and HiSeq_31, respectively). When the MinION metagenome was polished with the HiSeq reads (i.e. the MinION_RD metagenome), the query length:hit length of the contigs increased to ~ 1. These ratios indicate that while indels still impact the quality and accuracy of MinION sequences, this effect is strongly diminished when paired with HiSeq assembly (i.e. hybrid metagenome and Hybrid_5) or HiSeq polishing (i.e. MinION_RD metagenome).

Hybrid_5: A potential novel species of Octadecabacter. We recovered one high-quality and five medium-quality bins[39,49] from the hybrid assembly method and six medium-quality bins[49] from the HiSeq assembly method (Table 3). The most complete and least contaminated bins belonged to *Octadecabacter* (Hybrid_5, HiSeq_31), *Pseudomonas* (Hybrid_20), and *Polaromonas* (HiSeq_14). *Pseudomonas* and *Polaromonas* are familiar taxa in polar environments (e.g. active layer permafrost, sea ice, seawater)[51–53], and our MAGs matched
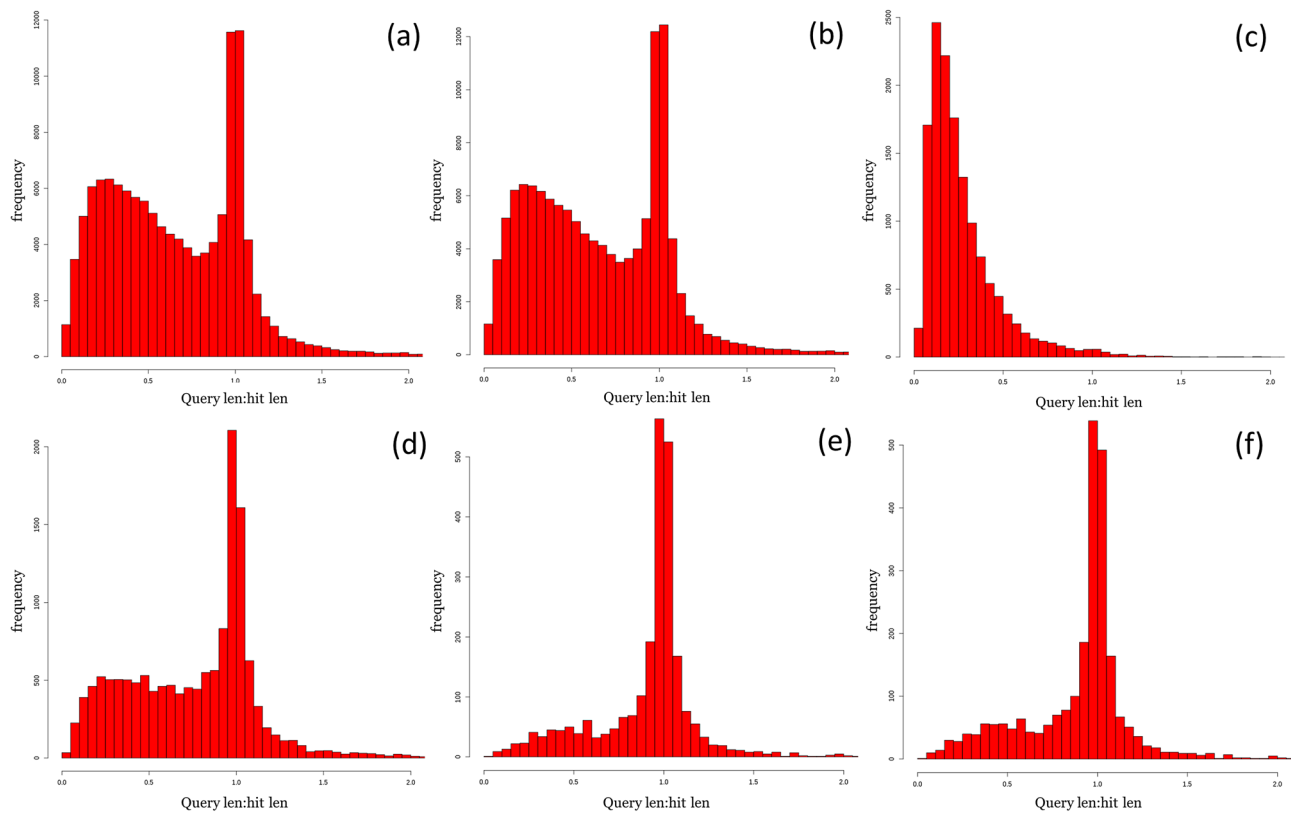
Figure 3. IDEEL plots of frequency (y-axis) vs. query length:hit length (x-axis) for: (**a**) hybrid metagenome; (**b**) HiSeq metagenome; (**c**) MinION metagenome; (**d**) polished MinION metagenome (MinION_RD); (**e**) Hybrid_5 MAG; (**f**) HiSeq_31 MAG. Query length:hit length is the ratio between the contig and its best match in the UniProt TREMBL database. If the contig contains few indels, this ratio will be ~ 1.

closely with *Pseudomonas fluorescens* (Hybrid_20) and *Polaromonas* sp. JS666 (HiSeq_14), although both MAGs lack a 16S gene. Hybrid_20 is a non-photosynthetic heterotroph capable of nitrate reduction and denitrification, consistent with the metabolic properties of *Pseudomonas fluorescens* and other polar pseudomonads[53,54]. HiSeq_14 is a non-photosynthetic heterotroph capable of sulfur oxidation and nitrate reduction, in line with its closest match *Polaromonas* sp. JS666 and other Arctic *Polaromonas* strains[51,52]. We have thus chosen to focus our analyses on the *Octadecabacter* MAG Hybrid_5, as the survival strategies and detailed metabolic characteristics of this genus are not as well-described as both *Pseudomonas* and *Polaromonas*. Additionally, Hybrid_5 exhibits unique features as compared with other members of *Octadecabacter* and some genomic features lacking in HiSeq_31, including a 16S gene.

Hybrid_5 has the highest completeness (90.9%) and lowest contamination (0.6%) of all MAGs produced in any assembly method (Table 3). It has a genome size of 3.27 Mbp, a mean contig length of 14.5 kbp, an N50 of 19.9 kbp, and was identified as belonging to the *Octadecabacter* genus by both GTDB and MiGA. The genus *Octadecabacter* is typified by *Octadecabacter arcticus* and *Octadecabacter antarcticus*, marine psychrophiles with a bipolar distribution and rich in octadecanoic acid[55,56]. One complete 16S rRNA gene was found in Hybrid_5 aligning to *O. arcticus* with 98% identity. When mapped to the *O. arcticus* genome, Hybrid_5 and its HiSeq counterpart HiSeq_31 show a reduced genome size, although Hybrid_5 traverses slightly more of the *O. arcticus* genome (Fig. 4). *O. arcticus* has a genome size of 5.2 Mbp[56], while Hybrid_5 and HiSeq_31 have genome sizes of 3.28 and 3.22 Mbp, respectively. This reduced size is likely due to the incompleteness of draft MAGs[57]. The average nucleotide identity (ANI) of Hybrid_5 was calculated at 93.51% with *O. arcticus*, below the threshold of 95% similarity for identical species, indicating that Hybrid_5 represents a potential novel species in the *Octadecabacter* genus.

Like *O. arcticus*, Hybrid_5 possesses genes encoding a complete tricarboxylic acid (TCA) cycle, Embden-Meyerhof glycolysis, oxidative phosphorylation, and the acetyl-CoA pathway (Fig. 5). A full Entner-Doudoroff pathway and glyoxylate shunt are also present, cycles that typically act as alternatives to glycolysis and the citric acid cycle, respectively[58,59]. The glyoxylate shunt may also allow Hybrid_5 to accumulate C4 compounds alongside the TCA cycle[60]. Hybrid_5 may be motile as it encodes for several flagellum production genes, including flagellar biosynthesis genes *flhAB* and *fliPQR*, basal body rod genes *flgCFG*, and motor switch genes *fliN* and *fliY*[61]. Features indicating potential anaerobic metabolic capabilities are also present, including full lactic acid and ethanol fermentation pathways, an incomplete Calvin-Benson-Bassham (CBB) cycle (glyceraldehyde-3P to ribulose-5P), an incomplete reductive TCA cycle (S-malate to citrate), and a complete acetyl-CoA pathway for potential carbon fixation and acetyl-CoA generation[62]. While Hybrid_5 possesses a full pyruvate dehydrogenase
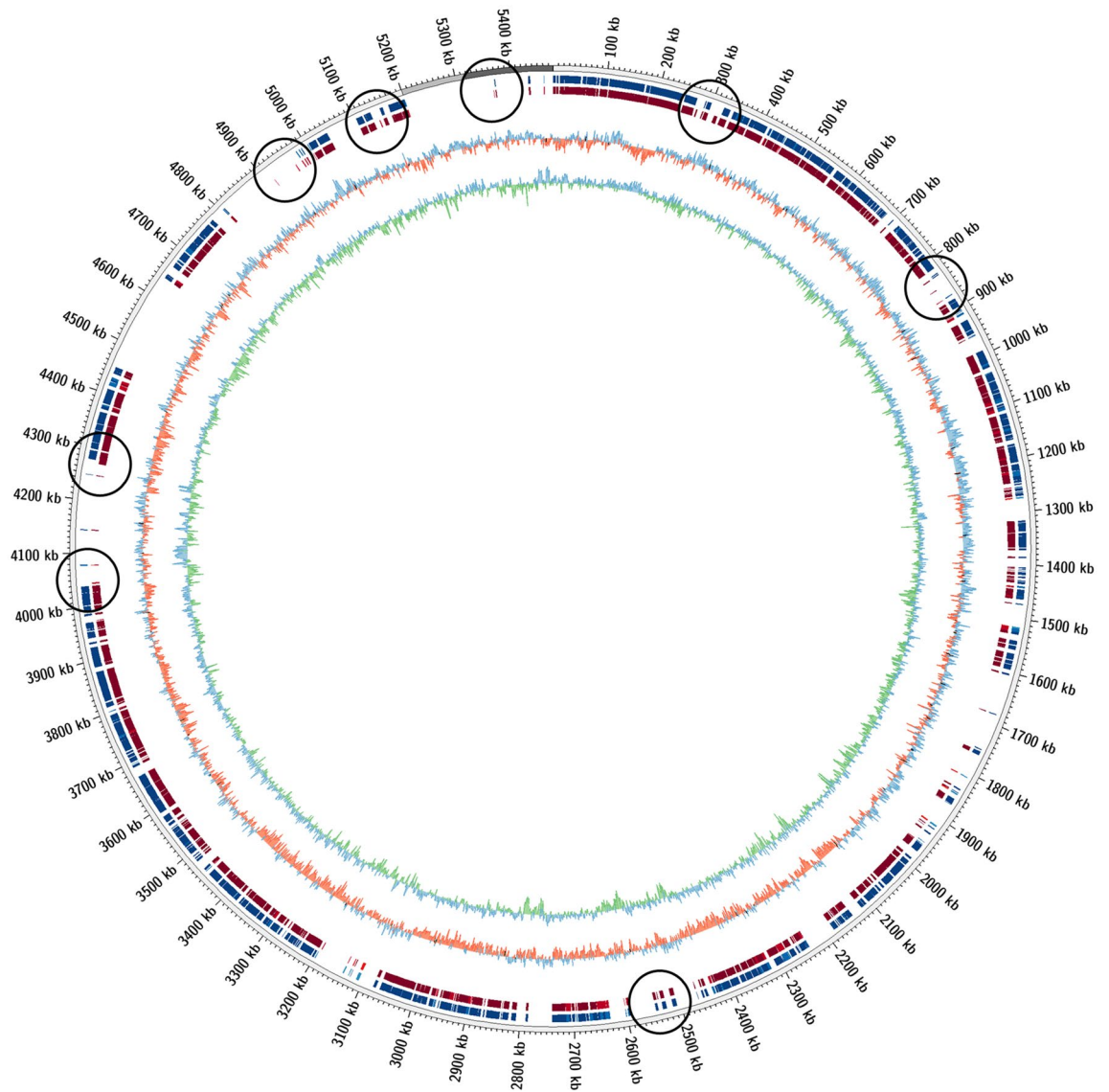
Figure 4. mummer2circos alignment of hybrid_5 (red) and HiSeq_31 (blue) to the *Octadecabacter arcticus* genome (grey). Center plots are GC skew (blue/red) and GC content/variation (blue/green). Areas where Hybrid_5 covers more of the *O. Arcticus* genome are circled in black.

complex [EC 1.2.4.1; EC 2.3.1.12; EC 1.8.1.4] for pyruvate oxidation under aerobic conditions, its genome also has pyruvate:ferredoxin oxidoreductase [EC 1.7.2.1] for anaerobic oxidation of pyruvate[63].

Hybrid_5 encodes a complete assimilatory sulfate reduction pathway (sulfate adenylyltransferase, *sat* and *cysND* [EC:2.7.7.4]; adenylsulfate kinase, *cysC* [EC:2.7.1.25]; phosphoadenosine phosphosulfate reductase, *cysH* [EC:1.8.4.8]; sulfite reductase (NADPH), *cysJI* [EC:1.8.1.2] and sulfite reductase (ferredoxin), *sir* [EC:1.8.7.1]) and a partial dissimilatory sulfate reduction pathway (sulfate adenylyltransferase, *sat* [EC:2.7.7.4]; adenylylsulfate reductase, *aprAB* [EC:1.8.99.2]). Hybrid_5 also contains genes for full pathways of dissimilatory nitrate reduction (nitrate reductase, *narGHI* [EC:1.7.5.1] and *napAB* [EC:1.9.6.1]; nitrite reductase, *nirBD* [EC:1.7.1.15]), assimilatory nitrate reduction (ferredoxin-nitrate reductase, *narB* [EC:1.7.7.2] and *nasAB* [EC:1.7.99.-]; ferredoxin-nitrite reductase, *nirA* [EC:1.7.1.1]) and denitrification (nitrate reductase, *narGHI* [EC:1.7.5.1] and *napAB* [EC:1.9.6.1]; NO-forming nitrite reductase, *nirK* [EC:1.7.2.1]; nitric oxide reductase, *norBC* [EC:1.7.2.5]; nitrous oxide reductase, *nosZ* [EC:1.7.2.4]). Nitrite oxidoreductase, *nxrAB* (EC:1.7.99.-), is present but Hybrid_5 lacks a complete nitrification pathway. *nar*, *nap*, and *nas* nitrate reductases are all encoded by Hybrid_5; Nap is located in the periplasmic membrane for dissipation of reducing power, Nar is a respiratory transmembrane protein that generates a proton motive force for ATP production, and Nas biosynthesizes N products in the cytoplasm[64,65], signifying that Hybrid_5 is able to use nitrate for redox balancing, as a terminal electron acceptor, and as a nitrogen source[65].

Many stress response and cold adaptation genes are present in Hybrid_5 (Supplementary Table 2), including general stress response (SOS-response transcriptional regulation), as well as more specialized genetic responses to extreme environments. Allen Bay experiences 24-h light from the months of May to August[66] which, coupled with the high albedo on sea ice, suggests that cryoconite consortia in this region must cope with deleterious
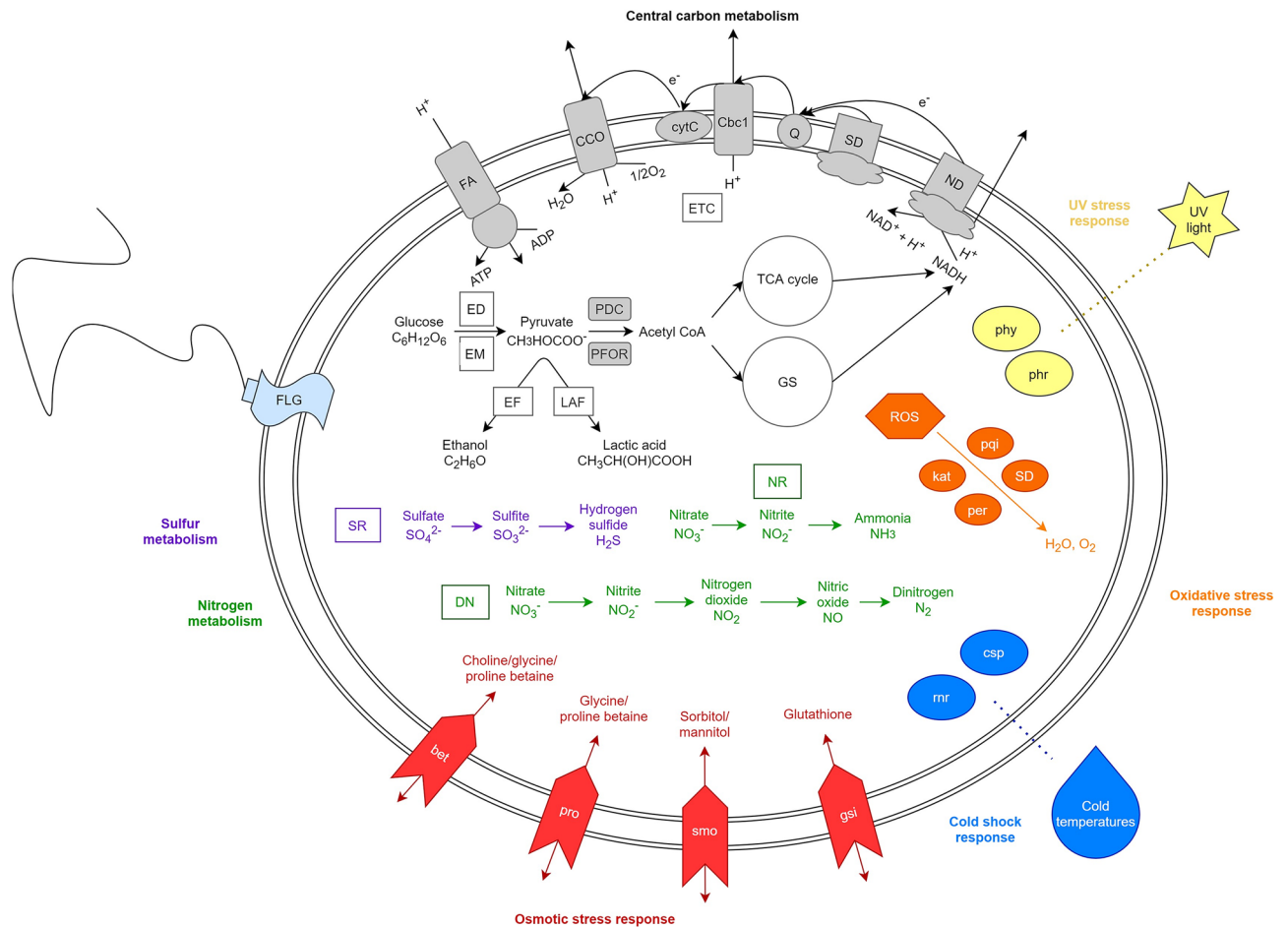
Figure 5. Model of Hybrid_5 cellular systems based on genomic data. Central carbon, nitrogen, and sulfur metabolisms are present, as well as stress response genes and pathways. Central carbon metabolism: Cbc1 = cytochrome bc1 complex; CCO = cytochrome c oxidase, cbb3-type; cytC = cytochrome C; ED = Entner-Doudoroff pathway; EF = ethanol fermentation; EM = Embden-Meyerhof glycolysis; ETC = electron transport chain; FA = F-type ATPase; Q = ubiquinone/quinone pool; GS = glyoxylate shunt; LAF = lactic acid fermentation; ND = NADH dehydrogenase; PDC = pyruvate dehydrogenase complex; PFOR = pyruvate:ferredoxin oxidoreductase; SD = succinate dehydrogenase; TCA = tricarboxylic acid. Nitrogen metabolism: DN = denitrification; NR = nitrate reduction. Sulfur metabolism: SR = sulfur reduction. Cold shock response: csp = cold shock protein; rnr = cold shock-induced ribonuclease R. Osmotic stress response: bet = glycine/ choline/proline betaine transporters; gsi = glutathione transporters; pro = glycine/proline betaine transporters; smo = sorbitol/mannitol transporters. Oxidative stress response: kat = catalase-peroxidase; per = peroxidase; pqi = paraquat-inducible protein A; ROS = reactive oxygen species; SD = superoxide dismutase; UV stress response: phr = photolyase; phy = phytoene synthase. FLG = flagellar biosynthesis genes.

UV radiation during these months[7,67]. Hybrid_5 contains genes for carotenoid production (phytoene synthase) and photolyase *phrB* for DNA repair caused specifically by UV damage[8,68,69]. Nutrient deprivation and rapid fluctuations are also common in sea ice and cryoconite environments[7,70], and Hybrid_5 appears to endure these difficulties with *pho* genes and glycerol-3-phosphate O-acyltransferase[68,69].

Cold temperatures such as those encountered in Allen Bay (1.33 °C; Table 1) require specialized genome-level adaptations, a number of which are present in Hybrid_5. These include cold shock proteins (*csp*), which bind to DNA and RNA to regulate transcription and translation at low temperatures[71], as well as molecular chaperones (e.g. *dnaJ, dnaK, hslO*) and chaperonins (e.g. *groES, groEL*) that ensure proper folding of cellular macromolecules[8,68,72]. Maintaining enzymatic rates for replication, transcription, and translation becomes more difficult at cold temperatures, and Hybrid_5 possesses several genes to counteract this reduced efficiency, such as recombination factors (e.g. *recA, recR*), DNA repair proteins (e.g. *recN, radA*), transcription termination factors (e.g. *rho, nusA*), and translation initiation factors (e.g. IF-1, IF-2)[69,72]. Proper protein folding is maintained in Hybrid_5 by peptidyl-prolyl cis–trans isomerase and *tig*[8]. Membrane fluidity is negatively impacted at cold temperatures, and Hybrid_5 appears to compensate for this effect with several membrane, peptidoglycan, and polysaccharide capsule alteration and production genes including peptidoglycan synthesis genes *murABCDEF*, fatty acid elongation genes *fabBFGH*, and exopolysaccharide biosynthesis gene *epsC*[68,69,73,74].

Osmotic and oxidative stress are also prevalent in cold environments due to increased environmental salinity and reactive oxygen species (ROS). The genomic adaptations present in Hybrid_5 to cope with rapid changes in environmental solute concentrations include several compatible solute (e.g. proline, glycine betaine, choline, carnitine, mannitol, sorbitol, glutamate) synthesis and transport genes, and transmembrane channel proteins (e.g. *kdpD*, NhaA family Na + :H + antiporter)[68,69,75]. Hybrid_5 also possesses a number of antioxidant defenses against oxidative stress, such as superoxide dismutase, glutathione synthesis and transport genes (e.g. *gshB*, *gsiABC*), peroxiredoxin, and catalase-peroxidase[68,69,72].

Discussion

The addition of MinION sequences to the HiSeq dataset resulted in a hybrid assembly superior to either of its constituent datasets. This is substantiated by the increase in contig length and classified coding sequences in the hybrid assembly, as well as the higher contig length, higher N50, higher completeness, and lower contamination in the hybrid MAGs. The increase of N50 increase in the hybrid MAGs indicates higher assembly contiguity than in the HiSeq dataset[36,39]. This is further supported by the general higher completeness of the hybrid MAGs and the higher number of classified coding sequences in the hybrid assembly, despite having fewer contigs than the HiSeq assembly. The hybrid assembly also had the greatest number of MAGs produced, indicating that the addition of even a small amount of MinION data (i.e. 1.2 Gbp and 22.9 Gbp of MinION and HiSeq data, respectively) increases output as well as quality. In terms of financial cost, a "starter pack" from ONT includes a MinION Mk1b device, a flow cell, and a sequencing kit for $1000 USD. One lane on a HiSeq 4000 typically costs ~ $2000–2500 USD, roughly double the price of the MinION starter park, and does not include the capacity to perform multiple runs in-house (although it produces sequences with ~ 100 × fewer errors).

One hybrid MAG (Hybrid_5) and three MinION MAGs (MinION_3, MinION_RD_2, MinION_RD_3) had full complements of ribosomal RNA (i.e. 5S, 23S, and 16S genes). Ribosomal RNA is notoriously difficult to recover from binning programs[76] but remains an important marker for microbial ecology community analyses, particularly the 16S rRNA gene[77]. The ability of nanopore sequencing to resolve repetitive regions, like 16S rRNA[35,50], likely contributed to the recovery of ribosomal RNA in MAGs containing MinION sequences and can allow for greater elucidation of SSU phylogeny than possible with HiSeq-only MAGs.). It is likely that with a larger nanopore dataset (e.g. metagenomes from high biomass environments, datasets generated by PromethION or multiple MinION sequencing runs), the benefits of adding long reads to short read datasets would be strengthened further. Indeed, it is possible to produce even genome-length contigs with nanopore sequencing[26,78,79], although these are typically from single isolates in culture or higher biomass environments than those used in this study.

All three assembly methods tested agreed in their general summary of dominant taxa present in the Allen Bay sea ice (i.e. Bacteroidetes, followed by Proteobacteria). The differences between the MinION and hybrid/HiSeq datasets further support the benefit of additional lab-based sequencing after initial field sequencing with the MinION. Given the improved average contig length and number of coding sequences classified by JGI IMG/M ER in the hybrid assembly, we have based our discussion of the microbial community on this dataset. The hybrid assembly also contains a slightly more diverse metagenome than the HiSeq, likely due to the addition of MinION long reads increasing classification. For example, the phylum Candidatus Gracilibacteria, an uncultured lineage with limited metabolisms[47], was detected exclusively in the hybrid and MinION metagenomes, and the hybrid dataset contained more unique genes in a higher abundance than the HiSeq dataset, demonstrating the value and utility of hybrid assembly for characterizing extreme astrobiology analogue environments. Studies of Canadian Arctic cryoconites are typically limited to glacial environments; to the best of our knowledge, the present study is the first to examine cryoconite microbial communities on sea ice in the Canadian Arctic.

The Allen Bay sea ice cryoconites are primarily comprised of Bacteria as Bacteroidetes (Flavobacteriales) and Proteobacteria (Burkholderiales, Pseudomonadales, and Rhodobacterales) (Fig. 2). Low amounts of Archaea were detected in our cryoconite metagenome, which may be due to the summer sampling time; archaeal abundances are known to increase in sea ice during winter[80]. This relatively low sequence diversity is common in sea ice and cryoconite communities, which often contain a few central taxa but exhibit high spatial and temporal variability between sites[7]. These prevalent taxa in our metagenome are ubiquitous in polar environments and are consistent with previous reports. Sea ice communities in the Canadian Arctic are frequently dominated by Bacteroidetes (*Flavobacterium*, *Polaribacter*), Alphaproteobacteria (SAR11, *Roseobacter*), and Gammaproteobacteria (*Moritella*)[80–83], although their abundance and activity vary seasonally.

Cyanobacteria and Proteobacteria often dominate in glacial and alpine cryoconites and these environments are strongly associated with high rates of primary production[2]. However, some glacial and alpine cryoconite environments can contain predominantly heterotrophic bacteria (e.g. Alphaproteobacteria, Betaproteobacteria, Bacteroidetes)[2,3,84]. Indeed, at lower latitudes and on smaller glaciers, heterotrophic dominance may be the norm, supported by allochthonous input of carbon[15]. Arctic cryoconites frequently contain high abundances of Proteobacteria (Alphaproteobacteria), Bacteroidetes, and Cyanobacteria, as well as eukaryotic algae, protists, and fungi[2,6,84–86]. Heterotrophic bacteria are abundant in sea ice, particularly first-year sea ice (e.g. Alphaproteobacteria, Gammaproteobacteria, Flavobacteria)[83], supporting the abundance of Bacteroidetes (Flavobacteria) and Betaproteobacteria (Burkholderiales) in our Allen Bay sea ice cryoconites.

The physico-chemical data of the Allen Bay sea ice cryoconites are presented in Table 1. The cryoconite water was cold (1.33 °C) and somewhat salty (0.13 ppt), with a pH of 7.96. While the average salinity of Arctic seawater is ~ 32.5–35 ppt[87,88], the salinity of nearby Resolute Bay's under-ice seawater has been previously reported as 0.2 ppt[87], a value consistent with the low salinity of 0.13 ppt noted here. The ammoniacal nitrogen present is higher than other Canadian Arctic sites[89]; like organic carbon, NH_4-N could also be being actively discharged from the surrounding sea ice or produced by the nitrate-reducing bacteria present. The dissolved oxygen of 12.90 mg/L is

similar to other cryoconites[90,91] and is high enough to support an aerobic community. Despite the low number of autotrophic community members, the oxygen content in the Allen Bay sea ice cryoconites is likely maintained with atmospheric exchange[92]. However, this contrasts with the negative oxidation–reduction potential[42] that indicates a reducing environment. Based on the microbial community and DO value, it is possible that the ORP probe was malfunctioning and the reported value is inaccurate.

The low amount of total organic carbon (TOC) present in our cryoconites (112.75 ppm) indicates an overall oligotrophic environment[93]. The TOC in the Allen Bay cryoconites is higher than reported values from seawater of nearby sites on Cornwallis Island (~ 0.8037 ppm)[83], but lower than other Arctic cryoconites ($> 10,000$ ppm)[93,94]. Our TOC values are more similar to those reported for Cornwallis Island sea ice (~ 5–217 ppm)[82,83], indicating that the carbon is likely exuded from the sea ice[8]. This low content of autotrophic taxa suggests that our cryoconites do not produce significant quantities of either autochthonous organic carbon or O_2. While cryoconites can be defined as occurring solely on glaciers, containing filamentous Cyanobacteria as a crucial structural component and a dominance of autotrophic community members[1,95], cryoconites have been previously described in non-glacial habitats and the microbial community of our samples is consistent with heterotrophic cryoconite communities[15]. Instead of being produced by filamentous Cyanobacteria, the particulate organic carbon in the Allen Bay sea ice cryoconites is likely provided by periodic allochthonous carbon input[3] and organic exudation from the surrounding sea ice[8]. The size of the Allen Bay sea ice cryoconites (< 3 cm) is also consistent with other cryoconites[2], rather than larger supraglacial or sea ice melt pools/ponds[96,97].

The most complete and contiguous MAG produced from any assembly method was Hybrid_5 (Table 3). Hybrid_5 also possesses more unique genes and more genes with higher copy numbers than its HiSeq counterpart, HiSeq_31, demonstrating the value of hybrid assembly in detailed studies of microbial ecology. Hybrid_5 was identified as a member of *Octadecabacter* by both GTDB and MiGA, with an average nucleotide identity (ANI) of 93.51% with *Octadecabacter arcticus*, implying that Hybrid_5 is a potential novel species in the marine psychrophilic *Octadecabacter* genus. The prevalence of *Octadecabacter* in marine environments suggests that Hybrid_5 is not a native cryoconite microorganism and entered this environment via the seawater.

Hybrid_5 contains genes consistent with an aerobic/microaerophilic and heterotrophic lifestyle (Fig. 5). *O. arcticus* shares these features, with the exception of the glyoxylate cycle; lacking isocitrate lyase [EC 4.1.3.1], *O. arcticus* instead uses the ethylmalonyl-CoA pathway as an alternative carbon metabolism[60]. The Hybrid_5 genome includes pathways for complete assimilatory sulfate reduction, assimilatory nitrate reduction, dissimilatory nitrate reduction, and denitrification, indicating that it can use sulfate or nitrate as a terminal electron acceptor. While denitrification is generally an anaerobic process, the presence of periplasmic *napAB* is a signpost of aerobic denitrification capabilities in many Proteobacteria, and suggests that Hybrid_5 is able to co-respire oxygen and nitrate[98]. Co-respiration of oxygen and nitrate is common in areas with rapidly fluctuating concentrations of oxygen, such as variable water flow within cryoconite holes[99].

Unlike Hybrid_5, *O. arcticus* possesses few anaerobic features in its genome. It lacks full fermentative pathways, sulfate reduction pathways, and nitrate reduction pathways. While *O. antarcticus* and *O. temperatus* are known to reduce nitrite only (via nitrite reductase, *nirBD* [EC:1.7.1.15])[56] and *O. sp. SW4* possesses a complete dissimilatory nitrate reduction pathway (nitrate reductase, *narGHI* [EC:1.7.5.1] and nitrite reductase, *nirBD* [EC:1.7.1.15])[100], no other species of *Octadecabacter* contains genes for full assimilatory nitrate reduction, denitrification, assimilatory sulfate reduction, lactic acid fermentation, or ethanol fermentation. The presence of these anaerobic features in the Hybrid_5 genome indicates the likelihood that it is a novel species of *Octadecabacter* that functions as a facultative anaerobe in its environment[101]. Hybrid_5 further differs from *O. arcticus* in its lack of xanthorhodopsin and gas vesicle formation genes, suggesting it does not use light-driven proton pumping as a source of energy production and is not buoyant in the cryoconite meltwater[56]. Hybrid_5 encodes for flagellar biosynthesis and may move via its flagella. However, although both *O. arcticus* and *O. antarcticus* possess flagellar gene clusters, they are non-motile and exact flagellar function remains to be explained[55,56].

Cryoconite holes expose their native microbes to numerous external stressors[1,7] and Hybrid_5 has a plethora of ways to cope with these extreme conditions (Supplementary Table 2), such as osmotic shock, reactive oxygen species (ROS), and freezing temperatures. The stress tolerance mechanisms of *Octadecabacter* species are largely unknown and the methods used by Hybrid_5 can elucidate these functions in this widespread polar marine genus. Hybrid_5 synthesizes and transports several compatible solutes, the accumulation of which prevents water loss without disrupting cellular function and reduces the intracellular freezing point[102]. In opposition, *O. arcticus* lacks an osmosensitive K + channel histidine kinase and sorbitol/mannitol transport system proteins, and synthesizes ectoine as an osmolyte[56]. As gas solubility increases at cold temperatures, so too does the concentration of ROS, necessitating that Hybrid_5 produce antioxidants to prevent cellular damage including superoxide dismutase, glutathione synthesis and transport genes, peroxiredoxin, and catalase-peroxidase[68,69,72], features it shares with *O. arcticus*.

Hybrid_5 also possesses adaptations within its central carbon metabolism that can contribute to oxidative stress tolerance. The glyoxylate shunt is up-regulated under oxidative stress, as it lacks the TCA cycle's decarboxylation steps that produce NADH[58,103], and glucose 6-phosphate dehydrogenase (*G6PD*/*zwf*) in the Entner-Doudoroff pathway converts NADP + to NADPH to protect cells from oxidative stress[59]. The pyruvate dehydrogenases *aceE* and *aceF* are involved in both oxidative and cold stress responses[72]. Cold temperatures reduce transcriptional and translational activity, protein folding rates, and membrane fluidity[68,69], and Hybrid_5 differs from HiSeq_31, the corresponding HiSeq-only MAG, in that its genome contains more of these protective features against cold temperatures and stresses. HiSeq_31 appears to lack *murB*, a transcription-repair coupling factor (superfamily II helicase), catalase peroxidase, an Na + :H + antiporter, and peroxidase, as well as fewer gene copies of cold shock protein, *cspA*.

The increase in quality generated by the addition of MinION sequences to HiSeq datasets also demonstrates the utility of hybrid assembly in astrobiology and biosignature detection studies; post-initial DNA detection with

the MinION, sample return and lab sequencing increase the information yielded to better characterize the natural consortia in extreme environments and elucidate further ways to detect them based on their genomes. MinION sequencing has strong potential for biosignature detection in future robotic and human planetary science missions based on its very small size/mass, minimal power requirements, and ability to produce reliable sequences from extreme environments[12,17,18]. Nucleic acids are complex organic polymers that can only be produced by living systems, thereby providing an unequivocal biosignature; coupled with a reliable database, the MinION could readily determine if DNA/RNA sequences produced are terrestrial contaminants[104], enabling a measure of protection against forward contamination in space missions and theoretically straightforward determination of non-terrestrial sequences (i.e. an unclassifiable, independent lineage). In this study, we have demonstrated the ability of the MinION to produce reliable metagenomes real-time in an extreme analogue environment, the Allen Bay sea ice cryoconites, which brings considerably more value than presence/absence DNA detection or single gene recognition (e.g. 16S); it allows for deeper exploration of phylogeny, as well as functional and metabolic potential. Analyses from samples returned to the laboratory can then be used to improve the characterization of the site's microbiology, as performed here with HiSeq sequencing and hybrid assembly, and inform on future studies for both biosignature detection and environmental microbiology of extreme environments.

Although considerable challenges remain in developing this technology for robotic space missions (e.g. automation of nucleic acid extraction and sequencing preparation, development of non-degrading solid state nanopores)[18], the MinION's proven functionality in microgravity and space conditions[17,19,20,105] indicate its suitability for future life detection missions. While we have shown that MinION sequences can be used to supplement and improve HiSeq data to produce superior MAGs[39], we have also demonstrated that the high sequencing error rate inherent in MinION technology currently precludes obtaining high-quality MinION-only MAGs. Significantly reducing the error rate through technology improvements[27] and/or metagenome-specific assembly and polishing pipelines will be crucial to bringing the MinION to a more robust technology readiness level (TRL) applicable to planetary science. The ideal MinION/nanopore sequencing technology would incorporate solid state nanopores to negate protein stability/degradation due to long flight times and radiation, be capable of detecting and sequencing very low concentrations of nucleic acids (DNA, RNA, xDNA), and generate low-error rate metagenomes and MAGs.

Conclusion

The present study combines short, accurate HiSeq reads with long, error-prone MinION sequences to produce more contiguous and more correct hybrid metagenomes and MAGs than either constituent dataset alone. MinION sequences generated in the Canadian high Arctic yielded a metagenome generally representative of the microbial community (> 50% Bacteroidetes), as well as taxa and metabolisms not detected by traditional short read sequencing (e.g. Candidatus Gracilibacteria). When used to supplement HiSeq sequencing data, the resulting hybrid metagenomes contained longer contigs and more classified coding sequences, and the hybrid MAGs had longer contigs, higher N50, higher completeness, and lower contamination than the HiSeq-only dataset. The increase in quality of the hybrid dataset is despite relatively low data output from the MinION and logistical restrictions of field sequencing. Additionally, none of the shortcomings of MinION sequencing were readily evident in the hybrid datasets (e.g. indel presence). We have also described a potential novel species of *Octadecabacter* (Hybrid_5) that conspicuously differs from its closest relative, *O. arcticus*, in its metabolic potential, possessing pathways for full nitrate reduction, denitrification, sulfate reduction, lactic acid fermentation, and ethanol fermentation pathways. Hybrid_5 likely functions as a facultative anaerobe in its environment and the Allen Bay sea ice cryoconite habitat is largely based on aerobic heterotrophy. This knowledge expands our knowledge of genome reconstruction with hybrid assembly in samples from extreme environments.

Methods

Sample site and collection. Samples were collected from Allen Bay sea ice (latitude: 74.44707; longitude: − 95.0348) near Resolute, Nunavut, Canada in the Canadian Arctic Archipelago on July 12, 2018 (Fig. 1). Allen Bay is located in a polar tundra; the area remains ice-covered for ~ 10 months per year[87], experiencing an average annual temperature of − 15.7 °C and precipitation of 161.2 mm[66]. Samples were collected aseptically directly on the sea ice from holes < 3 cm in diameter and all collection tools were sterilized with 70% ethanol. Latex gloves were worn during collection and samples were loaded into sterile falcon tubes. Samples were then transported from the collection site to the Polar Continental Shelf Program facility (PCSP) at the Martin Bergmann Complex in Resolute for immediate DNA extraction and MinION sequencing. Remaining samples were later transported to Montreal, Quebec, Canada at − 5 °C and subsequently stored at − 25 °C prior to analysis at McGill University (geochemical analyses) and Genome Quebec (HiSeq sequencing). DNA extraction was performed differently for MinION and Illumina sequencing and is described for each technique below.

Geochemical analyses. Dissolved oxygen, salinity, conductivity, total dissolved solids, barometric pressure, pH, and temperature were measured in situ with a YSI Pro2030 Field Dissolved Oxygen/Conductivity Meter (Cat No./ID:14–660-204). Nearby cryoconite samples were later analyzed for total organic carbon (TOC) and ammoniacal nitrogen (NH_4-N) at McGill University. TOC was measured by the ultra-violet /persulfate oxidation method on a Sievers Innovox TOC analyzer (General Electric Power and Water, Water and Process Technologies, Boulder, Colorado, USA). NH_4-N concentrations were determined spectrophotometrically using the modified indophenol blue method53 at 650 nm on a microplate reader (µQuant, BioTek Instruments, Winooski, Vermont, USA).

DNA extraction and MinION sequencing. In order to maximize differential coverage for binning[106], we performed multiple DNA extractions and MinION sequencing runs near Allen Bay at PCSP[41]. Extraction details are described in Supplementary Table 3. Cell lysis was performed using the SuperFastPrep-2 (MP Biomedicals), a handheld cell homogenizer. With 0.25–0.5 g of sample, we used Solution C1 from the Qiagen DNeasy Powerlyzer PowerSoil DNA kit (Cat No./ID: 12855-50) as well as either the Lysing Matrix A beads from MP Biomedicals (Cat No./ID: 116910050) or the PowerSoil beads supplied with the DNeasy kit for 45 s at the power setting of 25 for initial extraction. Following initial beating, one extract (Crude) was filtered with a 0.45 µm filter to remove large debris and treated no further. Two other extracts (C3FullM and VolTRAX) underwent the purification steps outlined in the DNeasy protocol (steps #5–19) and eluted in 50 µl of nuclease-free water. The final sample (C3Claremont) was extracted with the Claremont SimplePrep X1, an automated DNA extraction and purification system (Cat No./ID: 08.104.01), according to manufacturer's instructions using low inhibitor cartridges (Cat No./ID: 08.440.01) and 0.3–0.5 g of sample.

Four MinION sequencing runs were with these extracts as described in Supplementary Table 3. Three of the extracts (Crude, C3FullM, and C3Claremont) were prepared for sequencing using the Rapid PCR Barcoding kit (SQK-RPB004) and sequenced on separate R9.4 FLO-MIN106 flow cells with a MinION Mk1b device. One purified extract (VolTRAX) was prepared for sequencing with the VolTRAX sequencing kit (VSK-VSK002) and VolTRAX V2 device. It was subsequently sequenced on a R9.4 FLO-MIN106 flow cell with a MinION Mk1b device. All sequences were basecalled with MinKNOW version 1.7.7 and trimmed using Porechop version 0.2.3 (https://github.com/rrwick/Porechop).

DNA extraction and HiSeq sequencing. For HiSeq sequencing, a total of 10 extractions were performed in the laboratory at McGill University and pooled to maximize differential coverage for binning[106]. These extractions are described in Supplementary Table 4. Cryoconite 1 was extracted as follows: a crude extraction with the SuperFastPrep-2 (C1Crude), a full extraction with the SuperFastPrep-2 and the purification steps outlined in the DNeasy protocol (steps #5–19) (C1FullM), and a full and purified extraction with the DNeasy kit according to the manufacturer's instructions (C1Full). Cryoconite 2 was extracted as follows: a full extraction with the SuperFastPrep-2 and the purification steps outlined in the DNeasy protocol (steps #5–19) (C2FullM), and two full and purified extractions with the DNeasy kit according to the manufacturer's instructions (C2Full1 and C2Full2). Cryoconite 3 was extracted as follows: a full extraction with the SuperFastPrep-2 and the purification steps outlined in the DNeasy protocol (steps #5–19) (C3FullM), two full and purified extractions with the DNeasy kit according to the manufacturer's instructions (C3Full1 and C3Full2), and an extraction with the Claremont SimplePrep X1 (C3Claremont). The SuperFastPrep-2 and Claremont extractions were used as described for MinION sequencing in the previous section.

Samples were prepared for sequencing using the Nextera XT DNA Library Preparation Kit (Cat No./ID: FC-131–1096) and sequenced at Genome Quebec (Montreal, Canada) with an Illumina HiSeq 4000 (paired end 100 bp). All sequences were quality filtered with FastQC and trimmed with trimmomatic version 0.36.

Contig assembly and binning. To assess the value of adding MinION-generated data to traditional contig assembly and genome binning from metagenome methods, we performed three types of assembly and binning: HiSeq, hybrid, and MinION. HiSeq assembly and binning used contigs generated only from HiSeq sequencing. For HiSeq assembly, trimmed and quality filtered sequences were assembled into contigs with metaSPAdes version 3.13.0, a pipeline of the SPAdes assembler using default parameters. Contigs less than 500 bp were discarded and the total assembly length was 261,582,792 bp.

For HiSeq binning, Minimap2 version 2.13[107] and Samtools version 1.9[108] were used for mapping and sorting, respectively. Metabat version 2.12.1[109] was used to bin the assembled, mapped, and sorted contigs, followed by contamination reduction and completeness improvement with refinem version 0.0.25[110]. Final bin statistics were determined with CheckM version 1.0.13[111].

Hybrid assembly and binning used sequences from both HiSeq and MinION metagenomes, assembled together into hybrid contigs. For hybrid assembly, trimmed and quality filtered sequences from the HiSeq 4000 sequencer were assembled with the trimmed MinION sequences using the hybrid option of metaSPAdes version 3.13.0 (e.g. default parameters and inclusion of the MinION dataset with the "–nanopore" flag). Contigs less than 500 bp were discarded and the total assembly length was 276,349,668 bp.

For hybrid binning, Minimap2 version 2.13[107] and Samtools version 1.9[108] were used for mapping and sorting, respectively. Metabat version 2.12.1[109] was used to bin the assembled, mapped, and sorted contigs, followed by contamination reduction and completeness improvement with refinem version 0.0.25[110]. Final bin statistics were determined with CheckM version 1.0.13[111].

MinION assembly and binning used contigs generated only from MinION sequencing. For MinION assembly, sequences were corrected and assembled with Canu version 1.8[112] with parameters "corOutCoverage = 10,000," "corMinCoverage = 0," "corMhapSensitivity = high," and an assumed genome size of 4.25 Mbp, followed by polishing for consensus sequence improvement with Nanopolish version 0.12.0[78] and default parameters. Contigs less than 989 bp were discarded and the total assembly length was 25,193,561 bp.

For MinION binning, additional polishing using the HiSeq short reads with three rounds of Racon version 1.4.11[113] was also performed, as well as frameshift error correction with DIAMOND version 0.9.25[114] and MEGAN-ization to produce a final fasta file of the MinION contigs with MEGAN-lr[115]. Contigs with coverage significantly different from the mean were manually removed from select MinION bins in order to improve completeness and decrease contamination. Minimap2 version 2.13[107] and Samtools version 1.9[108] were used for mapping and sorting, respectively. Metabat version 2.12.1[109] was used to bin the assembled, mapped, and sorted

contigs, followed by contamination reduction and completeness improvement with refinem version 0.0.25[110]. Final bin statistics were determined with CheckM version 1.0.13[111].

Metagenomic contigs from all three datasets (HiSeq, hybrid, MinION) were uploaded to JGI IMG/M ER[116] for annotation. Bins from all three datasets (HiSeq, hybrid, MinION) were annotated with both RAST[117] and MetaErg[118], with taxonomy determination based on the Genome Taxonomy Database (GTDB)[100], and uploaded to the Microbial Genomes Atlas Online (MiGA)[119] for average nucleotide identity (ANI) determination. IDEEL[50] was also used to test for ORF interruptions in the contig sets and the completed MAGS using the UniProt TREMBL database[120]. Mummer2circos (https://github.com/metagenlab/mummer2circos) was used to plot hybrid and HiSeq MAGs against their closest taxonomic relative. Metagenome and MAG data are available in JGI (analysis project IDs Ga0450655, Ga0450656, Ga0450657) and GenBank (BioProject accession number PRJNA673486).

References

1. Cook, J., Edwards, A., Takeuchi, N. & Irvine-Fynn, T. Cryoconite: the dark biological secret of the cryosphere. *Prog. Phys. Geogr.* **40**, 66–111 (2016).
2. Edwards, A. *et al.* Coupled cryoconite ecosystem structure–function relationships are revealed by comparing bacterial communities in alpine and Arctic glaciers. *FEMS Microbiol. Ecol.* **89**, 222–237 (2014).
3. Edwards, A. *et al.* A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. *Environ. Res. Lett.* **8**, 035003 (2013).
4. Weisleitner, K., Perras, A., Moissl-Eichinger, C., Andersen, D. T. & Sattler, B. Source environments of the microbiome in perennially ice-covered Lake Untersee, Antarctica. *Front. Microbiol.* **10**, 1019 (2019).
5. Anesio, A. M., Lutz, S., Chrismas, N. A. & Benning, L. G. The microbiome of glaciers and ice sheets. *NPJ Biofilms Microbiomes* **3**, 1–11 (2017).
6. Mueller, D. R., Vincent, W. F., Pollard, W. H. & Fritsen, C. H. Glacial cryoconite ecosystems: a bipolar comparison of algal communities and habitats. *Nova Hedwig. Beih.* **123**, 173–198 (2001).
7. Maccario, L., Sanguino, L., Vogel, T. M. & Larose, C. Snow and ice ecosystems: not so extreme. *Res. Microbiol.* **166**, 782–795 (2015).
8. Edwards, A. *et al.* Microbial genomics amidst the Arctic crisis. *Microb. Genom.* mgen000375 (2020).
9. Fernández-Gómez, B. *et al.* Bacterial community structure in a sympagic habitat expanding with global warming: brackish ice brine at 85–90 N. *ISME J.* **13**, 316–333 (2019).
10. Wight, J., Varin, M.-P., Robertson, G. J., Huot, Y. & Lang, A. S. Microbiology in the field: construction and validation of a portable incubator for real-time quantification of coliforms and other bacteria. *Front. Public Health* **8**, 804 (2020).
11. Loman, N. J. & Watson, M. Successful test launch for nanopore sequencing. *Nat. Methods* **12**, 303–304 (2015).
12. Goordial, J. *et al.* In situ field sequencing and life detection in remote (79° 26′ N) Canadian High Arctic permafrost ice wedge microbial communities. *Front. Microbiol.* **8**, 2594 (2017).
13. Leggett, R. M. & Clark, M. D. A world of opportunities with nanopore sequencing. *J. Exp. Bot.* **68**, 5419–5429 (2017).
14. Carr, C. E. *et al.* in *Aerospace Conference, 2017 IEEE.* 1–18 (IEEE).
15. Edwards, A., Debbonaire, A. R., Sattler, B., Mur, L. A. & Hodson, A. J. Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N. *bioRxiv*, 073965 (2016).
16. Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y. & Tighe, S. W. Real-time DNA sequencing in the Antarctic dry valleys using the Oxford nanopore sequencer. *J. Biomol. Tech.: JBT* **28**, 2 (2017).
17. Castro-Wallace, S. L. *et al.* Nanopore DNA sequencing and genome assembly on the international space station. *Sci. Rep.* **7**, 18022 (2017).
18. Maggiori, C. *et al.* The limits, capabilities, and potential for life detection with MinION sequencing in a paleochannel mars analog. *Astrobiology* **20**, 375–393 (2020).
19. Carr, C. E. *et al.* Nanopore sequencing at Mars, Europa, and microgravity conditions. *NPJ Microgr* **6**, 1–6 (2020).
20. Sutton, M. A. *et al.* Radiation tolerance of Nanopore sequencing technology for life detection on Mars and Europa. *Sci. Rep.* **9**, 1–10 (2019).
21. Kaczmarek, Ł., Jakubowska, N., Celewicz-Gołdyn, S. & Zawierucha, K. The microorganisms of cryoconite holes (algae, Archaea, bacteria, cyanobacteria, fungi, and Protista): a review. *Polar Rec.* **52**, 176–203 (2016).
22. Zawierucha, K., Ostrowska, M. & Kolicka, M. Applicability of cryoconite consortia of microorganisms and glacier-dwelling animals in astrobiological studies. *Contemp. Trends Geosci.* **6**, 1–10 (2017).
23. Nordheim, T., Hand, K. & Paranicas, C. Preservation of potential biosignatures in the shallow subsurface of Europa. *Nat. Astron.* **2**, 673–679 (2018).
24. Pappalardo, R. *et al.* Science potential from a Europa lander. *Astrobiology* **13**, 740–773 (2013).
25. Goldstein, S., Beka, L., Graf, J. & Klassen, J. L. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genom.* **20**, 23 (2019).
26. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).
27. Jain, M. *et al.* MinION analysis and reference consortium: Phase 2 data release and analysis of R9. 0 chemistry. *F1000Research* **6**, 1–10 (2017).
28. Manley, L. J., Ma, D. & Levine, S. S. Monitoring error rates in Illumina sequencing. *J. Biomol. Tech: JBT* **27**, 125 (2016).
29. De Maio, N. *et al.* Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb. Genom.* **5**, 1–12 (2019).
30. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **3** (2017).
31. Schmidt, M.H.-W. *et al.* De novo assembly of a new Solanum pennellii accession using nanopore sequencing. *Plant Cell* **29**, 2336–2348 (2017).
32. Shin, S. C. *et al.* Nanopore sequencing reads improve assembly and gene annotation of the Parochlus steinenii genome. *Sci. Rep.* **9**, 1–10 (2019).
33. Tan, M. H. *et al.* Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. *GigaScience* **7**, gix137 (2018).
34. Gan, H. M. *et al.* Best foot forward: Nanopore long reads, hybrid meta-assembly and haplotig purging optimises the first genome assembly for the Southern Hemisphere blacklip abalone (Haliotis rubra). *Front. Genet.* **10**, 889 (2019).
35. Somerville, V. *et al.* Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* **19**, 143 (2019).

13

36. Sevim, V. *et al.* Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies. *Sci. Data* **6**, 1–9 (2019).
37. Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* **8**, giz43 (2019).
38. Bertrand, D. *et al.* Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937–944 (2019).
39. Overholt, W. A. *et al.* Inclusion of Oxford Nanopore long reads improves all microbial and phage metagenome-assembled genomes from a complex aquifer system. *Environ. Microbiol.* **22**, 4000–4013 (2020).
40. Brown, C. L. *et al.* Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. *Sci. Rep.* **11**, 1–12 (2021).
41. Raymond-Bouchard, I. *et al.* Assessment of automated nucleic acid extraction systems in combination with MinION sequencing as potential tools for the detection of microbial biosignatures. *Astrobiology* (2021) **Submitted (Submission No. AST-2020–2349)**.
42. Lipson, D. A., Jha, M., Raab, T. K. & Oechel, W. C. Reduction of iron (III) and humic substances plays a major role in anaerobic respiration in an Arctic peat soil. *J. Geophys. Res. Biogeosci.* **115**, 1–13 (2010).
43. Kirchman, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**, 91–100 (2002).
44. McCammon, S. A. & Bowman, J. P. Taxonomy of Antarctic Flavobacterium species: description of Flavobacterium gillisiae sp. nov., Flavobacterium tegetincola sp. nov., and Flavobacterium xanthum sp. nov., nom. rev. and reclassification of [Flavobacterium] salegens as Salegentibacter salegens gen. nov., comb. nov.. *Int. J. Syst. Evol. Microbiol.* **50**, 1055–1063 (2000).
45. Van Trappen, S., Vandecandelaere, I., Mergaert, J. & Swings, J. Flavobacterium degerlachei sp. nov., Flavobacterium frigoris sp. nov. and Flavobacterium micromati sp. nov., novel psychrophilic bacteria isolated from microbial mats in Antarctic lakes . *Int. J. Syst. Evol. Microbiol.* **54**, 85–92 (2004).
46. Van Etten, J., Graves, M., Müller, D., Boland, W. & Delaroque, N. Phycodnaviridae–large DNA algal viruses. *Adv. Virol.* **147**, 1479–1516 (2002).
47. Ley, R. E. *et al.* Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl. Environ. Microbiol.* **72**, 3685–3695 (2006).
48. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
49. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
50. Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953 (2019).
51. Margesin, R., Spröer, C., Zhang, D.-C. & Busse, H.-J. Polaromonasglacialis sp. nov. and Polaromonas cryoconiti sp. nov., isolated from alpine glacier cryoconite. *Int. J. Syst. Evol. Microbiol.* **62**, 2662–2668 (2012).
52. Mattes, T. E. *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.* **74**, 6405–6416 (2008).
53. Lin, X., Zhang, Z., Zhang, L. & Li, X. Complete genome sequence of a denitrifying bacterium, Pseudomonas sp CC6-YY-74, isolated from Arctic Ocean sediment. *Mar. Genom.* **35**, 47–49 (2017).
54. Ghiglione, J.-F., Richaume, A., Philippot, L. & Lensi, R. Relative involvement of nitrate and nitrite reduction in the competitiveness of Pseudomonas fluorescens in the rhizosphere of maize under non-limiting nitrate conditions. *FEMS Microbiol. Ecol.* **39**, 121–127 (2002).
55. Gosink, J., Herwig, R. & Staley, J. Octadecabacter arcticus gen. nov., sp. nov., and O. antarcticus, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water. *Syst. Appl. Microbiol.* **20**, 356–365 (1997).
56. Vollmers, J. *et al.* Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS ONE* **8**, e63422 (2013).
57. Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M. & Banfield, J. F. Accurate and complete genomes from metagenomes. *Genome Res.* **30**, 315–333 (2020).
58. Ahn, S., Jung, J., Jang, I.-A., Madsen, E. L. & Park, W. Role of glyoxylate shunt in oxidative stress response. *J. Biol. Chem.* **291**, 11928–11938 (2016).
59. DangThu, Q., Jang, S. H. & Lee, C. Biochemical comparison of two glucose 6-phosphate dehydrogenase isozymes from a cold-adapted Pseudomonas mandelii. *Extremophiles: Life under Extreme Cond.* **24**, 501–509 (2020).
60. Schneider, K. *et al.* The ethylmalonyl-CoA pathway is used in place of the glyoxylate cycle by Methylobacterium extorquens AM1 during growth on acetate. *J. Biol. Chem.* **287**, 757–766 (2012).
61. Liu, R. & Ochman, H. Origins of flagellar gene operons and secondary flagellar systems. *J. Bacteriol.* **189**, 7098–7104 (2007).
62. Oelgeschläger, E. & Rother, M. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch. Microbiol.* **190**, 257–269 (2008).
63. Hutcherson, J. A. *et al.* Amixicile, a novel strategy for targeting oral anaerobic pathogens. *Sci. Rep.* **7**, 1–14 (2017).
64. Kuypers, M. M., Marchant, H. K. & Kartal, B. The microbial nitrogen-cycling network. *Nat. Rev. Microbiol.* **16**, 263 (2018).
65. Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R. & Castillo, F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J. Bacteriol.* **181**, 6573–6584 (1999).
66. Canada, E. *Canadian Climate Normals 1981–2010 Station Data*, <https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=1776&autofwd=1> (2020).
67. Hatam, I. *et al.* Distinct bacterial assemblages reside at different depths in Arctic multiyear sea ice. *FEMS Microbiol. Ecol.* **90**, 115–125 (2014).
68. Goordial, J. *et al.* Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile Rhodococcus sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. *FEMS Microbiol. Ecol.* **92**, 1–11 (2016).
69. Raymond-Bouchard, I. *et al.* Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. *FEMS Microbiol. Ecol.* **94**, fiy023 (2018).
70. De Maayer, P., Anderson, D., Cary, C. & Cowan, D. A. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep.* **15**, 508–517 (2014).
71. Bakermans, C. *et al.* Genomic and expression analyses of cold-adapted microorganisms. *Polar Microbiol.: Life Deep Freeze* **1**, 126–155 (2011).
72. Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F. & Corbeil, J. Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Appl. Environ. Microbiol.* **78**, 549–559 (2012).
73. Barreteau, H. *et al.* Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 168–207 (2008).
74. Magnuson, E. *et al.* Thiomicrorhabdus streamers and sulfur cycling in perennial hypersaline cold springs in the Canadian high Arctic. *Environ. Microbiol.* **23**, 3384–3400 (2020).
75. Krulwich, T. A., Hicks, D. B. & Ito, M. Cation/proton antiporter complements of bacteria: why so large and diverse?. *Mol. Microbiol.* **74**, 257–260 (2009).
76. Alneberg, J. *et al.* Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome* **6**, 173 (2018).
77. Case, R. J. *et al.* Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl. Environ. Microbiol.* **73**, 278–288 (2007).

78. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **12**, 733–735 (2015).
79. Bian, L. *et al.* Chromosome-level genome assembly of the greenfin horse-faced filefish (Thamnaconus septentrionalis) using Oxford Nanopore PromethION sequencing and Hi-C technology. *Mol. Ecol. Resour.* **20**, 1069–1079 (2020).
80. Collins, R. E., Rocap, G. & Deming, J. W. Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ. Microbiol.* **12**, 1828–1841 (2010).
81. Alonso-Sáez, L., Sánchez, O., Gasol, J. M., Balagué, V. & Pedrós-Alio, C. Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ. Microbiol.* **10**, 2444–2454 (2008).
82. Garneau, M. -È. *et al.* Hydrocarbon biodegradation by Arctic sea-ice and sub-ice microbial communities during microcosm experiments, Northwest Passage (Nunavut, Canada). *FEMS Microbiol. Ecol.* **92**, fiw130 (2016).
83. Yergeau, E. *et al.* Metagenomic survey of the taxonomic and functional microbial communities of seawater and sea ice from the Canadian Arctic. *Sci. Rep.* **7**, 42242 (2017).
84. Edwards, A. *et al.* A distinctive fungal community inhabiting cryoconite holes on glaciers in Svalbard. *Fungal Ecol.* **6**, 168–176 (2013).
85. Edwards, A. *et al.* Possible interactions between bacterial diversity, microbial activity and supraglacial hydrology of cryoconite holes in Svalbard. *ISME J.* **5**, 150–160 (2011).
86. Edwards, A. *et al.* Contrasts between the cryoconite and ice-marginal bacterial communities of Svalbard glaciers. *Polar Res.* **32**, 19468 (2013).
87. Brown, K. A. *et al.* Inorganic carbon system dynamics in landfast Arctic sea ice during the early-melt period. *J. Geophys. Res.: Oceans* **120**, 3542–3566 (2015).
88. Brogi, S. R. *et al.* Optical and molecular characterization of dissolved organic matter (DOM) in the Arctic ice core and the underlying seawater (Cambridge Bay, Canada): Implication for increased autochthonous DOM during ice melting. *Sci. Total Environ.* **627**, 802–811 (2018).
89. Martin, J. *et al.* Prevalence, structure and properties of subsurface chlorophyll maxima in Canadian Arctic waters. *Mar. Ecol. Prog. Ser.* **412**, 69–84 (2010).
90. Bagshaw, E., Tranter, M., Wadham, J., Fountain, A. & Mowlem, M. High-resolution monitoring reveals dissolved oxygen dynamics in an Antarctic cryoconite hole. *Hydrol. Process.* **25**, 2868–2877 (2011).
91. Hodson, A. *et al.* The structure, biological activity and biogeochemistry of cryoconite aggregates upon an Arctic valley glacier: Longyearbreen, Svalbard. *J. Glaciol.* **56**, 349–362 (2010).
92. Falkner, K. K. *et al.* Dissolved oxygen extrema in the Arctic Ocean halocline from the North Pole to the Lincoln Sea. *Deep Sea Res. Part I* **52**, 1138–1154 (2005).
93. Singh, P., Hanada, Y., Singh, S. M. & Tsuda, S. Antifreeze protein activity in Arctic cryoconite bacteria. *FEMS Microbiol. Lett.* **351**, 14–22 (2014).
94. Kaštovská, K., Elster, J., Stibal, M. & Šantrůčková, H. Microbial assemblages in soil microbial succession after glacial retreat in Svalbard (High Arctic). *Microb. Ecol.* **50**, 396 (2005).
95. Segawa, T. *et al.* Biogeography of cryoconite forming cyanobacteria on polar and Asian glaciers. *J. Biogeogr.* **44**, 2849–2861 (2017).
96. Xu, D. *et al.* Contrasting community composition of active microbial eukaryotes in melt ponds and sea water of the Arctic Ocean revealed by high throughput sequencing. *Front. Microbiol.* **11**, 1170 (2020).
97. Trivedi, C. B., Lau, G. E., Grasby, S. E., Templeton, A. S. & Spear, J. R. Low-temperature sulfidic-ice microbial communities, Borup Fiord Pass, Canadian high Arctic. *Front. Microbiol.* **9**, 1622 (2018).
98. Ji, B. *et al.* Aerobic denitrification: a review of important advances of the last 30 years. *Biotechnol. Bioprocess Eng.* **20**, 643–651 (2015).
99. Marchant, H. K. *et al.* Denitrifying community in coastal sediments performs aerobic and anaerobic respiration simultaneously. *ISME J.* **11**, 1799–1812 (2017).
100. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
101. Simon, C., Wiezer, A., Strittmatter, A. W. & Daniel, R. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* **75**, 7519–7526 (2009).
102. Casanueva, A., Tuffin, M., Cary, C. & Cowan, D. A. Molecular adaptations to psychrophily: the impact of 'omic'technologies. *Trends Microbiol.* **18**, 374–381 (2010).
103. Van Acker, H. & Coenye, T. The role of reactive oxygen species in antibiotic-mediated killing of bacteria. *Trends Microbiol.* **25**, 456–466 (2017).
104. Fairén, A. G., Parro, V., Schulze-Makuch, D. & Whyte, L. Searching for life on Mars before it is too late. *Astrobiology* **17**, 962–970 (2017).
105. John, K. *et al.* The Biomolecule Sequencer Project: Nanopore sequencing as a dual-use tool for crew health and astrobiology investigations. In *Lunar and Planetary Science Conference* (2016).
106. Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).
107. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
108. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
109. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *Peer J* **3**, e1165 (2015).
110. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
111. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
112. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
113. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
114. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
115. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
116. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2011).
117. Aziz, R. K. *et al.* The RAST server: rapid annotations using subsystems technology. *BMC Genom.* **9**, 1–15 (2008).
118. Dong, X. & Strous, M. An integrated pipeline for annotation and visualization of metagenomic contigs. *Front. Genet.* **10**, 999 (2019).
119. Rodriguez-R, L. M. *et al.* The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* **46**, W282–W288 (2018).
120. Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

Acknowledgements

Author contributions

C.M. designed the experiment, collected samples, carried out DNA extractions and sequencing, bioinformatics analyses, and wrote the manuscript. I.R.B. designed the experiment, collected samples, carried out DNA extractions and sequencing, and contributed to manuscript writing. L.B. carried out DNA extractions. D.T. performed in situ geochemical analyses. L.W. oversaw the experiment. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00026-x.

Correspondence and requests for materials should be addressed to C.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.