Monitoring Microbial Populations and Antibiotic Resistance Gene Enrichment Associated with Arctic Waste Stabilization Ponds

Monica Gromala,a Josh D. Neufeld,a Brendan J. McConkeya

aDepartment of Biology, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT Wastewater management in the Canadian Arctic is challenging due to climate extremes, small population sizes, and lack of conventional infrastructure for wastewater treatment. Although many northern communities use waste stabilization ponds (WSPs) as their primary form of wastewater treatment, few studies have explored WSP microbial communities and assessed effluent impacts on receiving waters from a microbiological perspective. Here, we used 16S rRNA gene and metagenome sequencing to characterize WSP and receiving water microbial communities for two time points bracketing the spring WSP thaw in Baker Lake (Nunavut) and compared these results to other Nunavut WSPs in Cambridge Bay and Kugluktuk. Most amplicon sequence variants (ASVs) recovered from these WSP samples belonged to the phylum *Proteobacteria*, with considerable variation between the three locations and only six ASVs shared among the WSPs at >0.2% relative abundance. Wastewater indicator ASVs for the Baker Lake WSP were identified, and few indicator ASVs were detected in samples originating from other upstream or downstream sites. The metagenomic data revealed a strong enrichment of antibiotic resistance genes for WSP samples relative to downstream and reference samples, especially for genes associated with macrolide resistance. Together, our results provide a baseline characterization for WSP microbial communities, demonstrate how indicator ASVs can be used to monitor attenuation and dilution of effluent microorganisms, and reveal that WSPs can serve as hot spots for antibiotic resistance genes.

IMPORTANCE Given that the microbial communities of Arctic waste stabilization ponds (WSPs) are poorly studied to date, our characterization of multiple WSP systems and time points provides important baseline data that will assist with ongoing monitoring of effluent impacts on downstream aquatic ecosystems in the Arctic. This research also identifies indicator amplicon sequence variants (ASVs) of WSPs that will be helpful for future monitoring for WSP effluent attenuation and demonstrates that WSP microbial communities are enriched in antibiotic resistance genes. Given operational and infrastructure changes anticipated for wastewater treatment systems in the Arctic, baseline data such as these are essential for further development of safe and effective wastewater treatment systems.

KEYWORDS Arctic ecology, DNA sequencing, antibiotic resistance, microbial communities, waste stabilization ponds, wastewater

As a complex mixture of particulate and dissolved organic matter, nutrients, bacteria, metals, oils, and plastics, municipal wastewater effluent can have multiple impacts on the biological communities of receiving waters (1). When treated inadequately, municipal wastewater effluent can cause eutrophication due to the enrichment of macro- and micronutrients (2) and increased concentrations of nitrogen and phosphorus that promote the growth of harmful algal species (3). Such eutrophication

can lead to microbial proliferation, biomass production, oxygen depletion, and toxin release (4). Many studies have documented changes in microbial communities associated with effluent inputs, including phylum-level differences upstream and downstream of wastewater effluent discharge sites (5), increased microbial abundances at sites downstream of discharge sites (6), and downstream persistence of microorganisms from wastewater effluent (7). Additionally, microbial community diversity decreases were seen in downstream sediment (8), and impacted sediment sites were associated with decreased abundances of sulfate reducers, denitrifiers, and methanogens in comparison to sediments from upstream sites (9). Microplastics from wastewater effluent contribute to the formation of microbial biofilms (10), which increases their ability to persist downstream of effluent input.

Because anthropogenic antibiotics are processed through municipal wastewater, conventional wastewater treatment plant systems are recognized as hot spots for the dissemination and evolution of antibiotic resistance genes (ARGs) (11). Antibiotics in wastewater exert selective pressures on microbial communities in these systems (12) and enable further dissemination through horizontal gene transfer of ARG-containing plasmids that are passed from antibiotic-resistant bacteria (ARB) to sensitive recipients (13). Human pathogens may also be present within municipal wastewater, with the potential to acquire ARGs in this way.

Although wastewater treatment facilities have been constructed across temperate regions, including Canada (1), wastewater management in northern communities is limited due to the extreme climate of the Arctic and relatively small population sizes (14, 15). Isolated from the rest of Canada, northern communities also face logistical obstacles to the construction and operation of conventional wastewater treatment plants (WWTPs) (16). Arctic wastewater management solutions aim to prevent the deterioration of water quality and protect communities from exposure to potential pathogens while also minimizing operational and infrastructure costs (14). Currently, many northern Canadian communities use waste stabilization ponds (WSPs), also referred to as sewage lagoons or wastewater lagoons, for wastewater treatment (17, 18). In such small Nunavut communities, wastewater is transported directly from buildings of origin to the designated WSP by truck and, after deposit, freezes in the WSP for approximately 9 months of the year (19). During the other 3 months, which are collectively known as the "treatment season," higher temperatures and longer periods of sunlight result in a thawing of the waste and subsequent flow into receiving waters (18). Warmer conditions also enable microorganisms to perform aerobic and anaerobic degradation of organic matter found within the wastewater, thus reducing the amount of potentially harmful material entering downstream water bodies (20).

Although studies have profiled microbial communities of Arctic environments (21–27), microorganisms inhabiting Arctic WSPs and downstream bodies of water have not been investigated extensively. A study of a Finnish WSP identified dominant influent bacterial taxa related to the genera *Trichococcus*, *Methylorosula*, *Polaromonas*, and *Arcobacter*, as well as the families *Leptotrichiaceae*, *Comamonadaceae*, *Alcaligenaceae*, and *Holophagaceae* (28), although effluent samples from the wastewater treatment system were not investigated. Another study characterized the microbial communities of effluent water from Pond Inlet WSP in Nunavut, Canada (29), showing that the effluent contained dominant bacteria associated with the genera *Rhodoferax*, *Arcobacter*, and *Pseudomonas*.

Given that microbial communities of Arctic WSPs and their persistence in downstream receiving waters are poorly understood, the objectives of this study were to examine the WSP systems of Baker Lake, Cambridge Bay, and Kugluktuk, all in Nunavut, Canada, with a particular focus on sampling Baker Lake. The goal was to characterize microbial communities represented by the WSP samples, in comparison to those of upstream and downstream sites, and identify indicator taxa that may be monitored for downstream effluent contamination. Given the potential for WSPs to facilitate horizontal transfer of ARGs, we used metagenomics to profile Arctic WSP samples. We
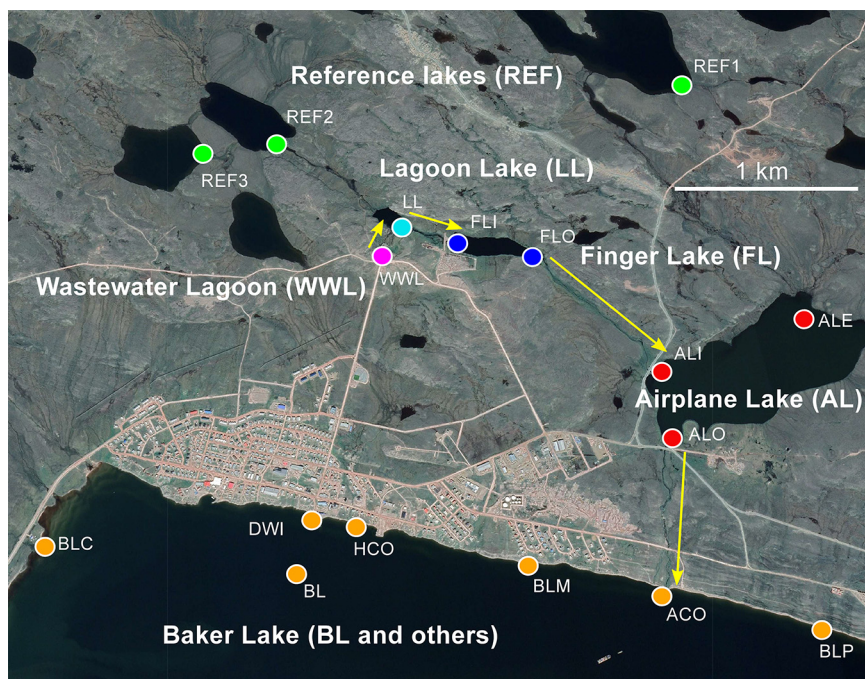
FIG 1 Map of sample sites for Baker Lake, Nunavut, Canada. Samples were grouped by lake as follows and as indicated by the colored circles: Baker Lake waste stabilization pond (WWL; magenta), Lagoon Lake (turquoise), Finger Lake (blue), Airplane Lake (red), Baker Lake (orange), and references (green). WWL, wastewater lagoon; LL, Lagoon Lake; FLI, Finger Lake inlet; FLO, Finger Lake outlet; ALI, Airplane Lake inlet; ALO, Airplane Lake outlet; ALE, Airplane Lake east; ACO, Airplane Creek outlet; BLP, Baker Lake port; BLM, Baker Lake mid; HCO, Hamlet Creek outlet; DWI, drinking water intake; BL, Baker Lake; BLC, Baker Lake Camp; REF1, reference 1; REF2, reference 2; REF3, reference 3. Flow of effluent from WSP is indicated by yellow arrows. The map was adapted from Google Earth.

anticipate that the work presented here will inform future efforts to increase WSP infrastructure as population sizes continue to grow in these northern communities.

RESULTS AND DISCUSSION

Temporal and spatial variability of Baker Lake WSP and downstream sites. Microbial samples were collected in the Baker Lake, vicinity during and after spring thaw on 13 to 16 July 2018 and 22 to 24 July 2018, as well as the Cambridge Bay WSP on 4 July and 25 July 2018 and the Kugluktuk WSP on 16 August 2018. The microbial community composition of water samples from the WSP and downstream receiving waters in Baker Lake (Nunavut; Fig. 1) were influenced by both temporal and spatial factors. Based on an ordination of 16S rRNA gene profiles (Fig. 2), samples collected during the first prethaw time point in Baker Lake (13 to 16 July 2018) separated from those collected during the second time point (22 to 24 July 2018). In addition, samples separated in ordination space based on distance from the WSP; sample sites that were located geographically closer to the WSP appeared closer to WSP samples in ordination space. Sites more distant from the WSP grouped with reference lake samples (Fig. 2). Temperature and other environmental data were also collected (see Table S1 in the supplemental material); however, these measurements did not exhibit a consistent direction of change between the two time points. Temperature, conductivity, dissolved oxygen (DO), and WSP-specific indicator taxa (indicator ASVs; see below) correlated with the separation of samples based on geographic location (Fig. 2).

Temporal and spatial variability were apparent within microbial community profiles when assessed at the phylum level (Fig. 3), revealing overall shifts with sampled distance from the WSP. The phylum *Proteobacteria* comprised >80% of the microbial community of the Baker Lake WSP at both time points. In contrast, changes in phylum abundance profiles occurred for the majority of sampled sites over the 10-day period
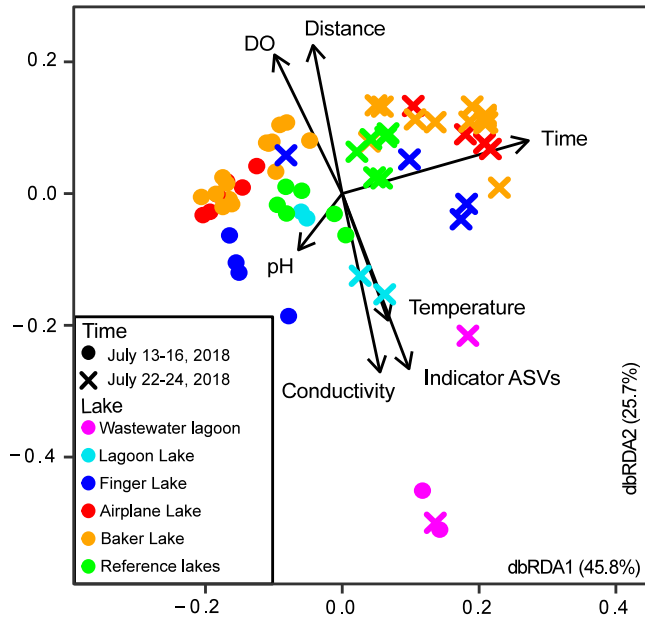
FIG 2 Distance-based redundancy analysis (db-RDA) based on principal-coordinate analysis results and environmental data for sites in Baker Lake, Nunavut, Canada.

between sampling time points. For example, although the phylum *Bacteroidetes* dominated the first time point samples collected from Airplane Lake and Baker Lake, higher levels of *Actinobacteria*, *Cyanobacteria*, and *Verrucomicrobia* were observed for the second time point samples of these lakes. Finger Lake samples were dominated by *Proteobacteria* for the first time point but then had higher levels of *Actinobacteria* and *Bacteroidetes* later in July 2018. The WSP and Lagoon Lake samples were relatively consistent temporally, with the exception of the 2WWLB ("second time point wastewater lagoon, replicate B") sample, which contained a high relative abundance of *Firmicutes*.

Uniqueness of Baker Lake WSP samples. The microbial community composition of the Baker Lake WSP samples was dominated by ASVs that affiliated with *Acinetobacter* and *Pseudomonas*, which together made up ~75% of the 16S rRNA gene profiles (Fig. 4). In general, few ASVs at >0.2% abundance were common between upstream reference sites, downstream sites, and other WSP sites. For example, for the first time point, samples from the first receiving body of water, Lagoon Lake, did not contain the ASVs associated with these two genera. Although the second time point samples from Lagoon Lake contain one of the ASVs affiliated with *Acinetobacter* and one affiliated with *Pseudomonas*, these ASVs are found at low abundances. Lagoon Lake also receives water from upstream lakes and is expected to contain microbial inputs from both the WSP and freshwater sources. However, the microorganisms detected in the WSP did not persist at high relative abundance in Lagoon Lake, suggesting attenuation by dilution or competition by microorganisms that are naturally present in the receiving waters.

Finger Lake, Airplane Lake, Baker Lake, and upstream reference lakes also contained very few ASVs at >0.2% abundance in common with the WSP (Fig. 4). Several samples from these lakes had no ASVs in common with the WSP at >0.2% abundance and were therefore not displayed. For the first time point, only 8 ASVs were shared between these lake samples and the WSP samples, which increased to 14 ASVs for the second time point, albeit at <1% abundance. Five of the shared ASVs were detected in upstream reference lake samples at >0.2% abundance, indicating that these microorganisms may be naturally present in the environment. An ASV associated with *Janthinobacterium* was detected at up to ~35% abundance in an upstream reference lake sample at the first time point but decreased to 2% at the second time point.
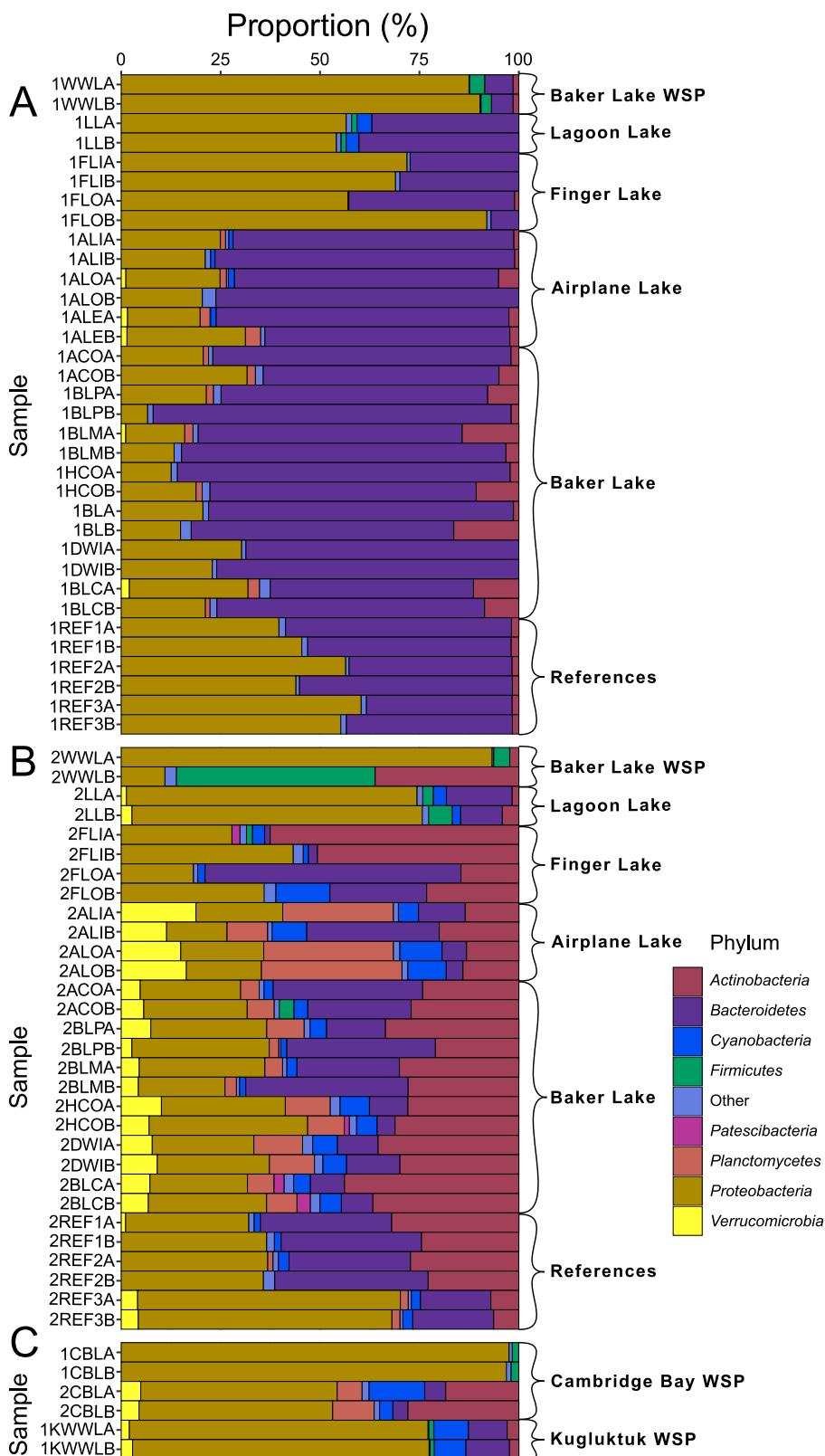
FIG 3 Relative abundance of phyla present at each site in Baker Lake, Cambridge Bay, and Kugluktuk, Nunavut, Canada. (A) Baker Lake, 13 to 16 July 2018; (B) Baker Lake, 22 to 24 July 2018. (C) Cambridge Bay WSP, 4 and 25 July 2018, and Kugluktuk WSP, 18 August 2018. Baker Lake sample names are coded by sites in Fig. 1, with prefixes 1 and 2 indicating sampling time and suffixes A and B indicating sample replicate. Cambridge Bay and Kugluktuk samples are labeled as CBL and KWWL, respectively.

Several strains of *Janthinobacterium* have been previously identified as psychrotolerant and psychrotrophic (30, 31), which would be consistent with their persistence at high abundance while water temperatures were relatively low.

Despite having similar microbial community profiles to the Baker Lake WSP at the phylum level (Fig. 3), samples from the WSPs in Cambridge Bay and Kugluktuk had very few ASVs at >0.2% abundance in common with the Baker Lake WSP (Fig. 4). The only ASV shared by all three WSPs was related to *Trichococcus*, which was also previously found in influent wastewater in an Arctic region in Finland (28). The Kugluktuk WSP samples shared ASVs associated with the *Acidovorax* and *Rhodoferax* genera. Both of these genera were found at up to ~15% abundance and were previously detected in samples from municipal WWTPs (32–34). The Cambridge Bay WSP first time point samples shared an ASV associated with *Psychrobacter* with the Baker Lake WSP samples. This ASV was detected at an abundance of ~83% in the Cambridge Bay WSP. Although *Psychrobacter* species are rare in wastewater environments, they are also generally psychrophilic (35), which may have allowed them to dominate the Cambridge Bay WSP due to the extreme cold. This ASV was not detected in the second time point Cambridge Bay WSP samples, which were collected following WSP discharge. Decreased WSP depth was likely associated with a temperature increase, which may have stimulated the growth of other microorganisms that displaced the *Psychrobacter* representatives.

Consistent wastewater indicator taxa affiliated with waste stabilization basin samples. An indicator species analysis (36) identified 46 highly specific wastewater indicator ASVs with an indicator value of >0.9 and a *P* value of <0.05 for the Baker Lake WSP and Lagoon Lake samples. Of these indicator ASVs, 13 were detected at >0.2% abundance in the Baker Lake WSP samples (Fig. 4). This includes the ASVs associated with *Acidovorax*, *Brevundimonas*, *Psychrobacter*, and *Trichococcus*, which were also detected in WSP samples from Cambridge Bay and/or Kugluktuk at >0.2% abundance. The ASV associated with the genus *Trichococcus* was the only indicator ASV that was detected in samples from all three WSPs and Lagoon Lake at >0.2% abundance. This ASV was also detected in a single sample (2ACOB; "second time point Airplane Creek Outlet, replicate B") at >0.2% abundance. Members of the *Trichococcus* genus are common wastewater pathogens previously identified in Arctic wastewater and tundra soils (28). Thus, it is possible that nutrient-rich conditions of the wastewater allowed members of this genus to persist in the Arctic WSPs, albeit at low abundances.

Only four of the indicator ASVs were detected in samples from multiple sites downstream of Lagoon Lake with >10 reads per sample (Fig. 5), including ASVs associated with *Acinetobacter*, *Ettlia oleoabundans*, *Pseudomonas*, and *Psychrobacter*. All other indicator ASVs were not detected in any downstream sites other than one sample (2ACOB), which was likely contaminated during sample processing because it contained 16 out of 46 indicator ASVs, whereas its replicate (2ACOA) only contained 1 indicator ASV. The ASV associated with *Acinetobacter* was also detected in reference lake samples during the second time point with >10 reads but only one sample (2REF1B) had >0.2% abundance. Although several species within the *Acinetobacter* genus are human pathogens, they are also relatively ubiquitous bacteria (37). This ASV cannot be linked directly with wastewater because it was also found at reference sites, which should have no wastewater input. The nutrient-rich environment of the WSP and psychrotolerant adaptations of the species itself may have enhanced the ability of this environmental organism to grow and persist in the Baker Lake WSP at >50% abundance.

Metagenomic sequence analysis of WSP samples. A subset of sample sites was used for metagenome sequencing, including each of the three WSPs, Lagoon Lake, the outlet sites of Finger Lake and Airplane Lake, the drinking water intake (DWI) site in Baker Lake, and reference lakes 1 and 2. All sampling sites were sequenced in duplicate at all available time points. A single representative site for each lake was selected, with reference lakes 1 and 2 used because their microbial community profiles were most similar to Baker Lake. Raw reads were analyzed using a hidden Markov model (HMM) approach for the identification of taxa based on the *rpoB* gene. Reads were
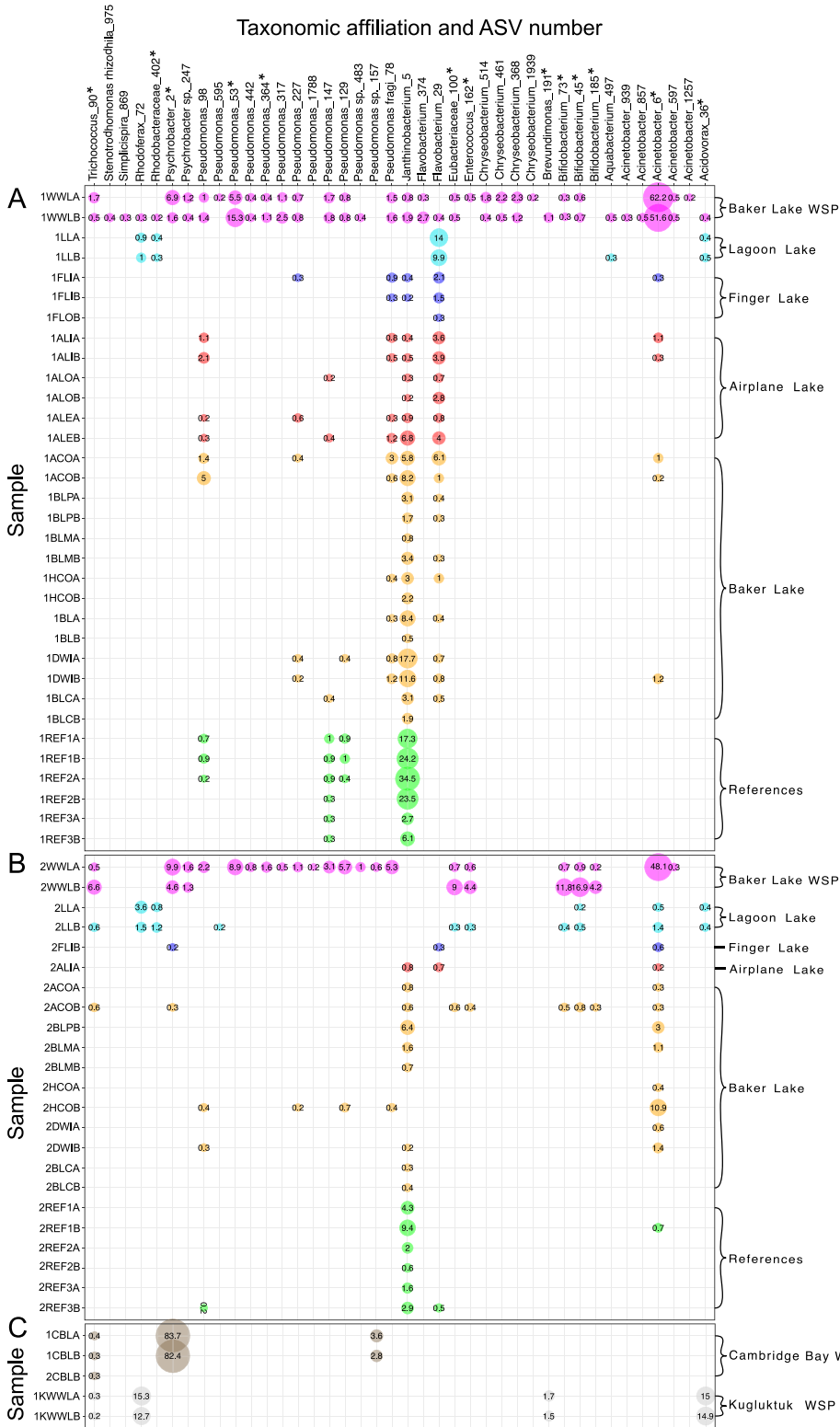
FIG 4 Relative abundance of amplicon sequence variants (ASVs) present at >0.2% abundance at each site in Baker Lake, Cambridge Bay, and Kugluktuk, Nunavut, Canada. (A) Baker Lake, 13 to 16 July 2018. (B) Baker Lake, 22 to 24 July 2018. (C) Cambridge Bay and Kugluktuk. ASVs with <0.2% abundance in 1WWLA, 1WWLB, and 2WWLA were not included. Samples containing >0.2% abundance of any of these ASVs were not included. Baker Lake sample names are coded by sites on Fig. 1, with prefixes 1 and 2 indicating sampling time and suffixes A and B indicating sample replicate. Cambridge Bay and Kugluktuk samples are labeled as CBL and KWWL, respectively. Indicator species are indicated with an asterisk.

also assembled into metagenome-assembled genomes (MAGs), and taxonomy was assigned to each MAG. These two data sets, along with the 16S rRNA gene profiles were compared.

Similar proportions of phyla for the same samples occurred in each of the three data sets (Fig. S1). This pattern was less apparent in the MAG taxonomic assignments, presumably because many metagenomic reads were not assigned to bins and were considered "unclassified" (Fig. S1). However, similar proportions of the phyla were still observed when comparing MAG taxonomic assignments to the other two data sets. *Proteobacteria* was the dominant phylum in the Baker Lake WSP across all three data sets. Both *Bacteroidetes* and *Actinobacteria* were highly abundant in the 16S rRNA gene and *rpoB* gene taxonomic assignments but were underrepresented in the MAG taxonomic assignments, likely due to low proportions of mapped reads. Based on *rpoB* gene taxonomic classifications, the *Actinobacteria* phylum was more prominent than suggested by the 16S rRNA gene data set. For example, in the first Baker Lake time point, the *Actinobacteria* had a relative abundance of ~5% in sample 1ALOA ("first time point Airplane Lake Outlet, replicate A") based on the 16S rRNA gene (Fig. S1A). However, this relative abundance increased to ~30% in the same sample based on the *rpoB* genes (Fig. S1B).

A Mantel test using distance matrices for all samples from each data set determined that the *rpoB* data set and 16S rRNA gene data set were most similar, producing statistically significant observed correlations (Table S2). The MAG data set had weaker correlations with the *rpoB* and 16S rRNA gene data sets, but they were still considered significantly positive correlations. Again, this is likely due to the underrepresentation of taxa in the MAG data set as a result of fewer reads mapping to bins. Hierarchical clustering of samples for each data set produced similar clusters among different data sets (Fig. S2). The 16S rRNA gene data set had clustering of samples most similar to the grouping of samples in the distance-based redundancy analysis (db-RDA) (Fig. 2). Replicate samples clustered together consistently. The Baker Lake WSP and Cambridge Bay first time point samples formed a separate cluster from other samples (Fig. S2). Based on hierarchical clustering, Baker Lake and reference lake samples were similar, with no distinct separation between the groups of samples. Lagoon Lake samples clustered with Cambridge Bay second time point samples and Kugluktuk samples, indicating that their microbial compositions were also similar to wastewater.

Antibiotic resistance genes are enriched in WSP samples. Metagenome-assembled genomes were further analyzed for completeness of KEGG pathways. The functional profiles of the samples were broadly similar across all samples (Fig. 6). Wastewater samples formed two clusters, with the Baker Lake WWL and two Cambridge Bay WWL samples in one group and the Kugluktuk WWL, Cambridge Bay WWL, and Lagoon Lake samples in a second group. There were few consistent differences in gene cluster completeness between the two WSP clusters and downstream and reference lakes. Overall, the functional composition of all samples appears to be similar despite the differences in taxonomic composition.

Wastewater is often considered to be a hot spot for antibiotic resistance gene dissemination. Studies calculating absolute abundance of ARGs using quantitative PCR (qPCR) have demonstrated that effluent wastewater had higher levels of ARGs than sites upstream of WWTPs (38–41). Another study, by Fitzpatrick and Walsh (42), used metagenomic techniques to demonstrate that environmental water samples contained fewer ARGs than WWTP sludge samples. Although metagenomic techniques are limited by the bias of DNA extraction methods, availability of database sequences for comparison, and sequence coverage required for rare bacterial species, these methods are still able to identify the presence of antibiotic resistance genes and compare relative abundances between different microbiomes.

Despite the broad functional similarities observed across metagenomes from all samples (Fig. 6), there were several antibiotic resistance gene families with increased
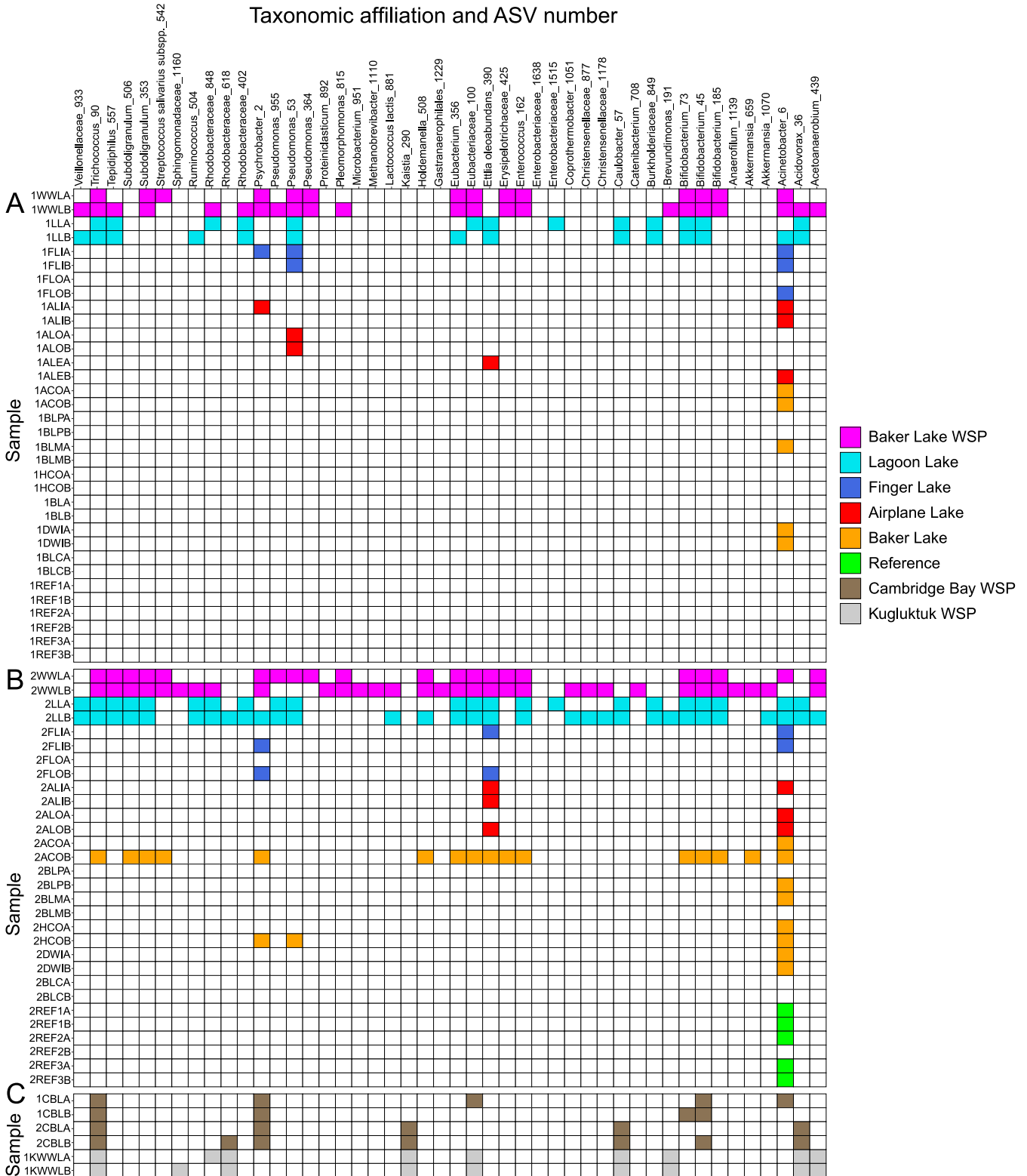
FIG 5 Presence/absence heatmap of wastewater indicator amplicon sequence variants (ASVs) at all sites in Baker Lake, Nunavut, Canada. (A) Baker Lake, 13 to 16 July 2018. (B) Baker Lake, 22 to 24 July 2018. (C) Cambridge Bay and Kugluktuk. Baker Lake sample names are coded by sites on Fig. 1, with prefixes 1 and 2 indicating sampling time and suffixes A and B indicating sample replicate. Cambridge Bay and Kugluktuk samples are labeled as CBL and KWWL, respectively. Boxes are filled in if the ASV had at least 10 reads for the site sample.

relative abundances in samples from the three WSPs and Lagoon Lake (Fig. 7). The gene families identified were related to macrolide resistance, class A beta-lactamases, ABC-F ribosomal protection, and rRNA methyl transferases. Within these four ARG classes, the gene families with highest abundance all appeared to be related to resistance to macrolide antibiotics, except for a single gene family (CARB-16) that was related to class A beta-lactamases. There were also several antibiotic resistance gene families identified in samples from the WSPs that were also identified in samples from sites downstream of the Baker Lake WSP and upstream reference sites. The genes within these families likely represent the naturally occurring environmental resistance genes.

Macrolide antibiotics are commonly used for the treatment of clinical infections, particularly infections of the upper respiratory tract and skin (43). Although speculative, it is possible that macrolide antibiotic usage by residents may be a source of low concentrations of antibiotics that accumulated in the WSPs over time. These resistance genes were not detected at high abundances in lakes downstream of Lagoon Lake in Baker Lake (Fig. 7), suggesting that the microorganisms carrying these ARGs do not persist in lakes further downstream. The identified ARG families related to macrolide resistance have not been linked with any particular species of bacteria; however, the ARG abundance correlates with the wastewater indicator ASVs identified in this study in that they were not detected in sites downstream of Lagoon Lake.

Conclusion. The frigid climate of the Canadian Arctic provides the need for specialized wastewater treatment solutions, often in the form of waste stabilization ponds. The Arctic WSPs studied here contained a microbial community composition dominated by members of the *Proteobacteria*. Despite having this similarity at the phylum level, each WSP had different dominant ASVs, which were also distinct from other Arctic freshwater lake samples. The Baker Lake WSP and downstream samples had considerable temporal and spatial variability, and the microbial community composition appeared to shift considerably during the 10-day period between sampling time points. In addition, samples from sites that were located geographically further from the WSP had less similar microbial community composition to the WSP than samples from sites located closer. Wastewater indicator ASVs were also identified for the Baker Lake WSP, and most of these ASVs were not found in samples from sites downstream of Lagoon Lake, suggesting that wastewater microorganisms are rapidly attenuated with increasing distance from the WSP in this system.

Antibiotic resistant gene families were quantified, and several of these gene families were found to be enriched in samples from the three WSPs. However, on a broad scale, functional profiles of these same appeared to be very similar to all of the other samples. In particular, ARG families related to macrolide resistance were highly abundant in the three WSPs, which may be related to the use of antibiotics within the local community. The abundant ARG families were not detected at high abundances in downstream and reference sites, suggesting that ARB from the WSP do not persist in lakes downstream of Lagoon Lake. Although specific ARBs were not identified in this study, it is possible that microorganisms containing ARGs include the wastewater indicator organisms, as these microorganisms were almost exclusively found in the Baker Lake WSP and Lagoon Lake. Overall, this study has provided a baseline characterization of microbial communities in Arctic WSPs, which will aid in the future development of a safe and effective wastewater treatment system for the Baker Lake community.

MATERIALS AND METHODS

Study locations. This study sampled the microbial communities of WSPs and receiving water in three communities in Nunavut, Canada, Baker Lake (64.3176°N, 96.0220°W), Cambridge Bay (69.1169°N, 105.0597°W), and Kugluktuk (67.8252°N, 115.0966°W).

Baker Lake is an inland community in Nunavut, located west of Hudson's Bay in the Kivalliq region. The Baker Lake WSP consists of a single WSP located approximately 1 km north of the community. Wastewater is collected and discharged into the WSP year-round, resulting in the accumulation of wastewater into an ice block over the 9-month period when temperatures are subzero. The ice typically
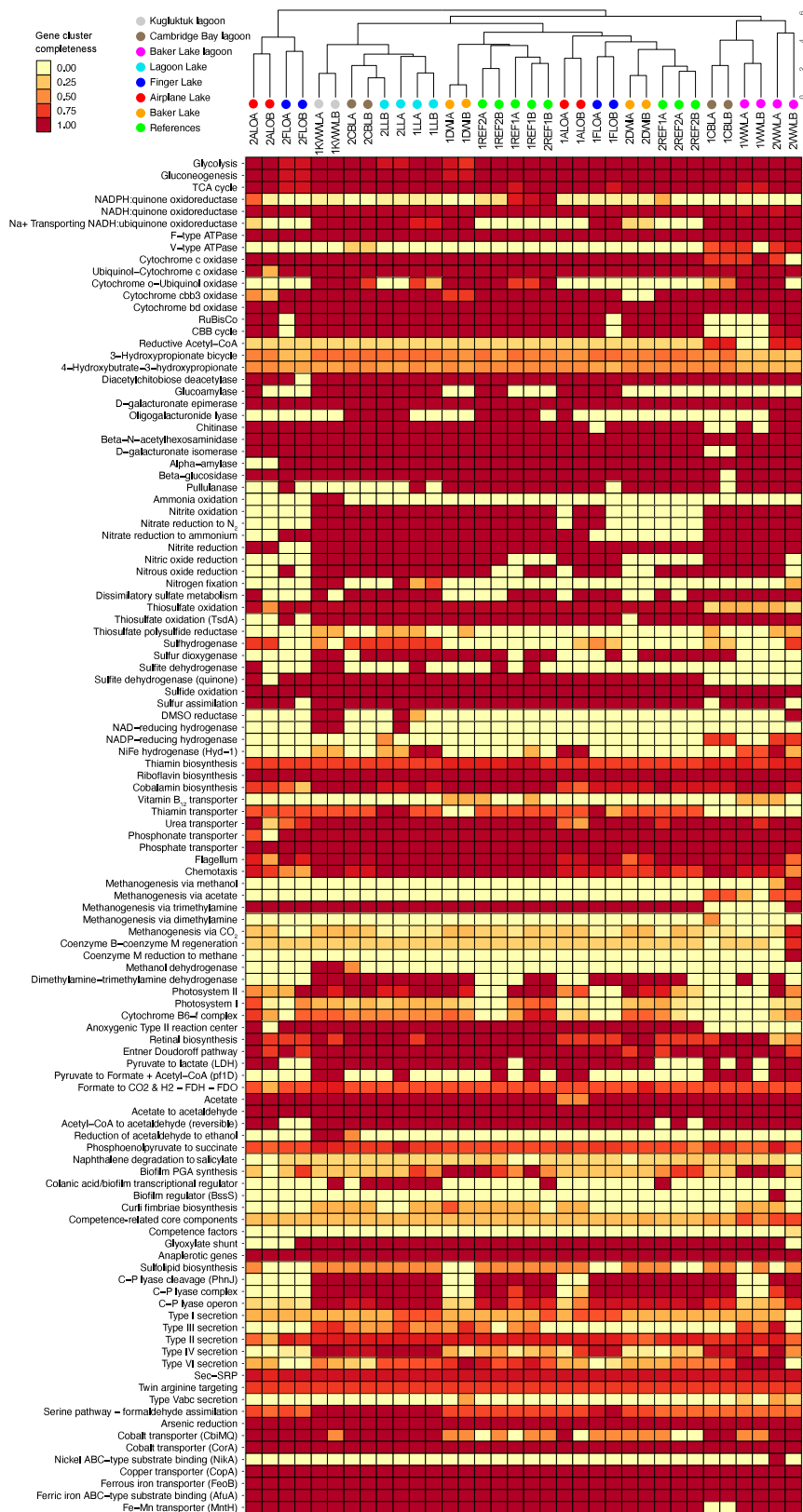
FIG 6 Heatmap of completeness of KEGG pathways based on predicted proteins from metagenome-assembled genomes (MAGs). Baker Lake sample names are coded by sites on Fig. 1, with prefixes 1 and 2 indicating sampling time and suffixes A and B indicating sample replicate. Cambridge Bay and Kugluktuk samples are labeled as CBL and KWWL, respectively.

FIG 7 Heatmap of log$_{10}$-transformed RPKM values of antibiotic resistance gene (ARG) families based on raw metagenomic reads. Baker Lake sample names are coded by sites on Fig. 1, with prefixes 1 and 2 indicating sampling time and suffixes A and B indicating sample replicate. Cambridge Bay and Kugluktuk sample names are labeled as CBL and KWWL, respectively. A dendrogram for sample similarity is shown at the top of the figure.
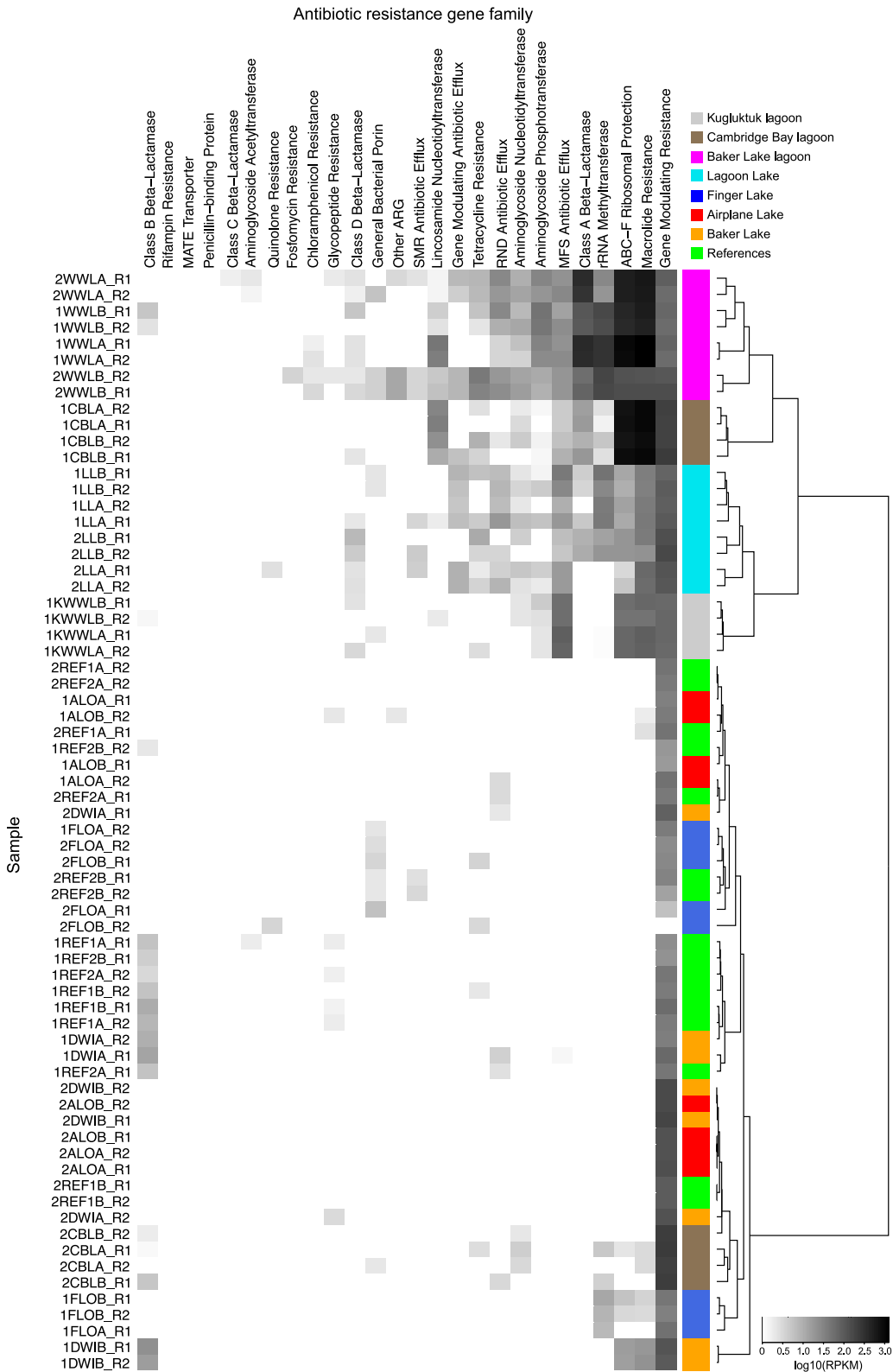
thaws in early June, allowing the water to flow into Lagoon Lake. From there, the water flows into Finger Lake, followed by Airplane Lake, before finally discharging into Baker Lake (Fig. 1). Each of these lakes, along with three upstream reference lakes were sampled at two time periods from 13 to 16 July 2018 and 22 to 24 July 2018, with the exception of the ALE and BL sites, which were only sampled within the first time period. The initial spring thaw occurred 2 to 3 weeks before the first set of samples was collected.

Cambridge Bay is located on Victoria Island in the Kitikmeot region of Nunavut. Wastewater is dumped into an engineered WSP, located approximately 1 km northeast of the community, where primary treatment occurs. The wastewater is discharged once a year every summer, when it passes through a natural tundra wetland before discharging into Cambridge Bay. Samples were taken from the WSP on 4 July and 25 July 2018.

Kugluktuk is located at the mouth of the Coppermine River in the Kitikmeot region of Nunavut. Wastewater is dumped into a single-cell, lined sewage lagoon. The wastewater is discharged once a year from the northwest corner of the lagoon. The wastewater then passes through a natural tundra wetland and flows into Coronation Gulf. Samples were taken from the WSP at a single time point on 16 August 2018.

Sample collection. Water samples were collected $\sim$4 to 5 m from the shoreline using a telescoping swing sampler and attached bottle to avoid disturbance of sediments. Samples were collected in duplicate from the top 10 cm of the water column. Water was filtered through a sterile 0.22-μm Sterivex-GV pressure filter unit (EMD Millipore Corporation, MA, USA) using a 60-ml Luer-Lok tip syringe (Becton, Dickinson, NJ, USA). Filtration was done until the filter became plugged or a maximum of 500 ml was filtered. Water was then expelled from the filter unit, and samples were kept in a cooler and surrounded by frozen ice packs until delivered to the lab. Once received by the lab, samples were stored at $-20°$C.

DNA extraction. Filters were removed aseptically from filter housing units and cut in half. One half was used for DNA extractions using the PowerSoil DNA isolation kit (Mo Bio, CA, USA) following the manufacturer's instructions, and the other half was frozen as a backup. PowerBead tubes were heated at 70°C for 10 min prior to bead beating with a FastPrep-24 device (MP Biomedicals, CA, USA) at 5.5 m/s for 45 s. Extraction kit negative controls were processed alongside samples.

16S rRNA gene amplicon library preparation and sequencing. Using a modification of a previously published amplification and sequencing method (44), the V4-V5 region of the 16S rRNA gene was amplified from genomic DNA extracts using primers 515F-Y/926R (45, 46). Samples were randomly assigned barcodes. Triplicate PCR products were pooled, and DNA for each sample was quantified using a NanoDrop 2000c spectrophotometer using absorbance at 260 nm. The DNA was then relatively quantified using gel electrophoresis. DNA quantification estimates were used to normalize the DNA so that approximately equal amounts of DNA from each sample were combined and sequenced in order to obtain a similar number of reads per sample. Nontemplate negative controls were added as 5 μl volumes, and positive controls, containing equal amounts of *Thermus aquaticus* and *Aliivibrio fischeri* were added at half the amount of the samples. The 16S rRNA gene amplicon library was purified and prepared for sequencing on a MiSeq device (Illumina, CA, USA) using the provided HT1 buffer to dilute the library to about 6 pM. Paired-end sequencing (2 $\times$ 250 bases) was performed on a MiSeq device, generating $\sim$18 million paired-end reads.

Taxonomic profiling and beta diversity analysis. Paired-end reads were demultiplexed using the MiSeq Reporter software. Subsequent FASTQ files were imported into QIIME 2 (47), and adapter and primer sequences were removed using Cutadapt (48). Reads were then trimmed, denoised, dereplicated, and merged using DADA2 (49) within QIIME 2. Reads were trimmed based on a minimum quality score of 25 for the nucleotide position. This produced a feature table containing amplicon sequence variants (ASVs). The April 2018 SILVA (50) release 132, 97% taxonomy classification for the 16S rRNA gene, was used to train the naive Bayes classifier, and taxonomy was assigned to sequence variants within the feature table. The feature table was collapsed to the phylum level, and the uncollapsed ASV table was rarefied to a sampling depth based on the sample with the lowest number of reads (Data Set S1). A beta diversity analysis was conducted with scripts contained within the QIIME 2 pipeline. Phylogenetic trees were created using the ASV table, and beta diversity analysis was performed using the same sampling depth as for rarefaction. Principal-coordinate analysis (PCoA) was performed using the weighted UniFrac distance metric. A distance-based redundancy analysis (ds-RDA) was performed using the PCoA results using the vegan package in R (51).

Indicator species analysis. Rarefied ASV tables exported from QIIME 2 were used for an indicator species analysis, using only Baker Lake WSP samples, Lagoon Lake samples, and reference lake samples. The Baker Lake WSP and Lagoon Lake samples were assigned to one group, and reference lake samples were assigned to another group. Indicator values for each ASV within these samples were calculated using the multipatt function from the indicspecies R package (52); 1,000 permutations were tested to determine statistical significance. The ASVs with an indicator value (IndVal) of at least 0.9 and P value of $<$ 0.05 were selected as indicator ASVs.

Metagenome sequencing. Genomic DNA was aliquoted for a subset of samples. Each sample location selected was sequenced in duplicate for all time points available. Library preparation, quality control, and whole-metagenome shotgun sequencing were performed by the Farncombe Metagenomics Facility at McMaster University on a HiSeq 1500 device (Illumina), generating a total of $\sim$500 million paired-end reads.

HMM taxonomic profiling. Raw metagenomic reads were analyzed using a development version of MetAnnotate (53) (available at https://github.com/MetAnnotate/MetAnnotate/tree/develop, commit ID:

6e92c0e). A hidden Markov model (HMM) for the gene for the beta subunit of bacterial RNA polymerase (*rpoB*) was downloaded from the TIGRFAMs database (54) and was used to profile the metagenomes by assigning taxonomy to reads identified as hits to the HMM. Forward and reverse reads were analyzed separately using HMM E values of ≤1e−3 and RefSeq HMM E values of ≤1e−6. Default values were used for all other parameters.

Metagenome assembly and binning. Metagenomic reads were processed using the ATLAS pipeline (release 2.0.6), which performs quality control, assembly, annotation, binning, and read mapping (55). Quality control was performed with the BBtools suite (56) tools to eliminate adapters, remove PCR duplicates, and then trim and filter reads based on quality scores and lengths of reads. Paired-end reads were assembled using the metaSPAdes (57) assembly tool. Prodigal (58) was used to predict n reading frames (ORFs) in contigs, and translated gene products were mapped to the eggNOG (59) catalogue using DIAMOND (60). Taxonomy was assigned using BAT (61), mapping genes to the GenBank protein database (62). Contigs were then binned using MetaBAT2 (63) and MaxBin2 (64), followed by DASTool (65) as a final binner to dereplicate, aggregate, and score the bins and create high-quality metagenome assembled genomes (MAGs). Because some of the same genomes were identified in multiple samples, resulting in multiple bins for the same MAG, dRep (66) was also used to select the best bin for each MAG.

Mantel test and hierarchical clustering. The ASV tables, *rpoB* tables, and MAG tables were compared to one another with a Mantel test to assess correlations between the three data sets. Bray-Curtis dissimilarity matrices were produced for each of the three tables in R using the vegdist function in the vegan package (51). The Mantel test was then performed on each pair of matrices using the mantel.rtest function from the ade4 R package (67), with 1,000 permutations to verify statistical significance. Hierarchical clustering was then performed on each distance matrix using the hclust function in the stats R package (68). The group average agglomeration method was used to calculate the distance between groups.

Functional profiling of assembled metagenomes and identification of ARGs. KofamScan (69) was used to assign KEGG ontology to the predicted proteins of MAGs. The output file from KofamScan was then parsed using KEGG-Decoder (70) to identify the completeness of each identified KEGG pathway in each sample. Forward and reverse reads from each set of paired-end metagenomic read files were analyzed separately to identify and quantify antibiotic resistance genes. This was done using the precomputed 2017 Antibiotic Resistance Factors marker collection from the ShortBRED (71) documentation pages (available at http://huttenhower.sph.harvard.edu/shortbred). Gene families were grouped manually into larger groups based on information from the Comprehensive Antibiotic Resistance Database (CARD) (72).

Data availability. Metagenome and 16S rRNA gene amplicon sequences are available in the Sequence Read Archive (SRA) at NCBI under BioProject accession number PRJNA600216.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.
SUPPLEMENTAL FILE 2, XLSX file, 7.5 MB.

REFERENCES

1. Chambers PA, Allard M, Walker SL, Marsalek J, Lawrence J, Servos M, Busnarda J, Munger KS, Adare K, Jefferson C, Kent RA, Wong MP. 1997. Impacts of municipal wastewater effluents on Canadian waters: a review. Water Qual Res J Canada 32:659–713. https://doi.org/10.2166/wqrj.1997.038.
2. Hu Z, Houweling D, Dold P. 2012. Biological nutrient removal in municipal wastewater treatment: new directions in sustainability. J Environ Eng 138:307–317. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000462.
3. Harper D. 1992. What is eutrophication?, p 1–28. *In* Eutrophication of freshwaters. Springer, The Netherlands.
4. Heisler J, Glibert PM, Burkholder JM, Anderson DM, Cochlan W, Dennison WC, Dortch Q, Gobler CJ, Heil CA, Humphries E, Lewitus A, Magnien R, Marshall HG, Sellner K, Stockwell DA, Stoecker DK, Suddleson M. 2008. Eutrophication and harmful algal blooms: a scientific consensus. Harmful Algae 8:3–13. https://doi.org/10.1016/j.hal.2008.08.006.
5. Wakelin SA, Colloff MJ, Kookana RS. 2008. Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. Appl Environ Microbiol 74:2659–2668. https://doi.org/10.1128/AEM.02348-07.

6. Cébron A, Coci M, Garnier J, Laanbroek HJ. 2004. Denaturing gradient gel electrophoretic analysis of ammonia-oxidizing bacterial community structure in the lower Seine River: impact of Paris wastewater effluents. Appl Environ Microbiol 70:6726–6737. https://doi.org/10.1128/AEM.70.11.6726-6737.2004.

7. Goi-Urriza M, Capdepuy M, Raymond N, Quentin C, Caumette P. 1999. Impact of an urban effluent on the bacterial community structure in the Arga River (Spain), with special reference to culturable Gram-negative rods. Can J Microbiol 45:826–832. https://doi.org/10.1139/w99-073.

8. Drury B, Rosi-Marshall E, Kelly JJ. 2013. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 79:1897–1905. https://doi.org/10.1128/AEM.03527-12.

9. Atashgahi S, Aydin R, Dimitrov MR, Sipkema D, Hamonts K, Lahti L, Maphosa F, Kruse T, Saccenti E, Springael D, Dejonghe W, Smidt H. 2015. Impact of a wastewater treatment plant on microbial community composition and function in a hyporheic zone of a eutrophic river. Sci Rep 5:17284. https://doi.org/10.1038/srep17284.

10. Eckert EM, Di Cesare A, Kettner MT, Arias-Andres M, Fontaneto D, Grossart H-P, Corno G. 2018. Microplastics increase impact of treated wastewater on freshwater microbial community. Environ Pollut 234:495–502. https://doi.org/10.1016/j.envpol.2017.11.070.

11. Li A-D, Metch JW, Wang Y, Garner E, Zhang AN, Riquelme MV, Vikesland PJ, Pruden A, Zhang T. 2018. Effects of sample preservation and DNA extraction on enumeration of antibiotic resistance genes in wastewater. FEMS Microbiol Ecol 94:1–11. https://doi.org/10.1093/femsec/fix189.

12. Zhang Y, Li A, Dai T, Li F, Xie H, Chen L, Wen D. 2018. Cell-free DNA: a neglected source for antibiotic resistance genes spreading from WWTPs. Environ Sci Technol 52:248–257. https://doi.org/10.1021/acs.est.7b04283.

13. Li A-D, Li L-G, Zhang T. 2015. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. Front Microbiol 6:1025. https://doi.org/10.3389/fmicb.2015.01025.

14. Daley K, Jamieson R, Rainham D, Truelstrup Hansen L. 2018. Wastewater treatment and public health in Nunavut: a microbial risk assessment framework for the Canadian Arctic. Environ Sci Pollut Res Int 25:32860–32872. https://doi.org/10.1007/s11356-017-8566-8.

15. Krkosek WH, Ragush C, Boutilier L, Sinclair A, Krumhansl K, Gagnon G, Jamieson RC, Lam B. 2012. Treatment performance of wastewater stabilization ponds in Canada's Far North. J Cold Regions Eng 26:612–622. https://doi.org/10.1061/9780784412473.061.

16. Chaves-Barquero LG, Luong KH, Mundy CJ, Knapp CW, Hanson ML, Wong CS. 2016. The release of wastewater contaminants in the Arctic: a case study from Cambridge Bay, Nunavut, Canada. Environ Pollut 218:542–550. https://doi.org/10.1016/j.envpol.2016.07.036.

17. Ragush CM, Schmidt JJ, Krkosek WH, Gagnon GA, Truelstrup Hansen L, Jamieson RC. 2015. Performance of municipal waste stabilization ponds in the Canadian Arctic. Ecol Eng 83:413–421. https://doi.org/10.1016/j.ecoleng.2015.07.008.

18. Huang Y, Truelstrup Hansen L, Ragush CM, Jamieson RC. 2018. Disinfection and removal of human pathogenic bacteria in Arctic waste stabilization ponds. Environ Sci Pollut Res 25:32881–32893. https://doi.org/10.1007/s11356-017-8816-9.

19. Neudorf KD, Huang YN, Ragush CM, Yost CK, Jamieson RC, Truelstrup Hansen L. 2017. Antibiotic resistance genes in municipal wastewater treatment systems and receiving waters in Arctic Canada. Sci Total Environ 598:1085–1094. https://doi.org/10.1016/j.scitotenv.2017.04.151.

20. Tilley E, Ulrich L, Lüthi C, Reymond P, Zurbrügg C. 2014. Compendium of sanitation systems and technologies, 2nd ed. Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Duebendorf, Switzerland.

21. Frank-Fahle BA, Yergeau É, Greer CW, Lantuit H, Wagner D. 2014. Microbial functional potential and community composition in permafrost-affected soils of the NW Canadian Arctic. PLoS One 9:e84761. https://doi.org/10.1371/journal.pone.0084761.

22. Hansen AA, Herbert RA, Mikkelsen K, Jensen LL, Kristoffersen T, Tiedje JM, Lomstein BA, Finster KW. 2007. Viability, diversity and composition of the bacterial community in a high Arctic permafrost soil from Spitsbergen, Northern Norway. Environ Microbiol 9:2870–2884. https://doi.org/10.1111/j.1462-2920.2007.01403.x.

23. Larose C, Dommergue A, Vogel TM. 2013. The dynamic Arctic snow pack: an unexplored environment for microbial diversity and activity. Biology (Basel) 2:317–330. https://doi.org/10.3390/biology2010317.

24. Nemergut DR, Costello EK, Meyer AF, Pescador MY, Weintraub MN, Schmidt SK. 2005. Structure and function of alpine and arctic soil microbial communities. Res Microbiol 156:775–784. https://doi.org/10.1016/j.resmic.2005.03.004.

25. Wilhelm RC, Niederberger TD, Greer C, Whyte LG. 2011. Microbial diversity of active layer and permafrost in an acidic wetland from the Canadian High Arctic. Can J Microbiol 57:303–315. https://doi.org/10.1139/w11-004.

26. Steven B, Pollard WH, Greer CW, Whyte LG. 2008. Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high Arctic. Environ Microbiol 10:3388–3403. https://doi.org/10.1111/j.1462-2920.2008.01746.x.

27. Steven B, Briggs G, McKay CP, Pollard WH, Greer CW, Whyte LG. 2007. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. FEMS Microbiol Ecol 59:513–523. https://doi.org/10.1111/j.1574-6941.2006.00247.x.

28. Gonzalez-Martinez A, Sihvonen M, Muñoz-Palazon B, Rodriguez-Sanchez A, Mikola A, Vahala R. 2018. Microbial ecology of full-scale wastewater treatment systems in the Polar Arctic Circle: archaea, bacteria and fungi. Sci Rep 8:2208. https://doi.org/10.1038/s41598-018-20633-5.

29. Huang Y. 2017. Characterization of microbial communities, disinfection and removal of human pathogenic bacteria in Arctic wastewater stabilization ponds. Doctoral dissertation. Dalhousie University, Halifax, Nova Scotia, Canada.

30. Gong X, Skrivergaard S, Korsgaard BS, Schreiber L, Marshall IPG, Finster K, Schramm A. 2017. High quality draft genome sequence of *Janthinobacterium psychrotolerans* sp. nov., isolated from a frozen freshwater pond. Stand Genomic Sci 12:8. https://doi.org/10.1186/s40793-017-0230-x.

31. Schloss PD, Allen HK, Klimowicz AK, Mlot C, Gross JA, Savengsuksa S, McEllin J, Clardy J, Ruess RW, Handelsman J. 2010. Psychrotrophic strain of *Janthinobacterium lividum* from a cold Alaskan soil produces prodigiosin. DNA Cell Biol 29:533–541. https://doi.org/10.1089/dna.2010.1020.

32. Heylen K, Lebbe L, De Vos P. 2008. *Acidovorax caeni* sp. nov., a denitrifying species with genetically diverse isolates from activated sludge. Int J Syst Evol Microbiol 58:73–77. https://doi.org/10.1099/ijs.0.65387-0.

33. Schulze R, Spring S, Amann R, Huber I, Ludwig W, Schleifer K-H, Kämpfer P. 1999. Genotypic diversity of *Acidovorax* strains isolated from activated sludge and description of *Acidovorax defluvii* sp. nov. Syst Appl Microbiol 22:205–214. https://doi.org/10.1016/S0723-2020(99)80067-8.

34. McIlroy SJ, Starnawska A, Starnawski P, Saunders AM, Nierychlo M, Nielsen PH, Nielsen JL. 2016. Identification of active denitrifiers in full-scale nutrient removal wastewater treatment systems. Environ Microbiol 18:50–64. https://doi.org/10.1111/1462-2920.12614.

35. Bowman JP. 2006. The genus *Psychrobacter*, p 920–930. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds), The prokaryotes. Springer, New York, NY.

36. Dufrêne M, Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr 67:345–366. https://doi.org/10.2307/2963459.

37. Al Atrouni A, Joly-Guillou M-L, Hamze M, Kempf M. 2016. Reservoirs of Non-*baumannii Acinetobacter* species. Front Microbiol 7:49. https://doi.org/10.3389/fmicb.2016.00049.

38. Hayward JL, Jackson AJ, Yost CK, Truelstrup Hansen L, Jamieson RC. 2018. Fate of antibiotic resistance genes in two Arctic tundra wetlands impacted by municipal wastewater. Sci Total Environ 642:1415–1428. https://doi.org/10.1016/j.scitotenv.2018.06.083.

39. Khan FA, Söderquist B, Jass J. 2019. Prevalence and diversity of antibiotic resistance genes in Swedish aquatic environments impacted by household and hospital wastewater. Front Microbiol 10:688. https://doi.org/10.3389/fmicb.2019.00688.

40. Ng C, Tay M, Tan B, Le T-H, Haller L, Chen H, Koh TH, Barkham TMS, Thompson JR, Gin KY-H. 2017. Characterization of metagenomes in urban aquatic compartments reveals high prevalence of clinically relevant antibiotic resistance genes in wastewaters. Front Microbiol 8:2200. https://doi.org/10.3389/fmicb.2017.02200.

41. Cacace D, Fatta-Kassinos D, Manaia CM, Cytryn E, Kreuzinger N, Rizzo L, Karaolia P, Schwartz T, Alexander J, Merlin C, Garelick H, Schmitt H, de Vries D, Schwermer CU, Meric S, Ozkal CB, Pons MN, Kneis D, Berendonk TU. 2019. Antibiotic resistance genes in treated wastewater and in the receiving water bodies: a pan-European survey of urban settings. Water Res 162:320–330. https://doi.org/10.1016/j.watres.2019.06.039.

42. Fitzpatrick D, Walsh F. 2016. Antibiotic resistance genes across a wide variety of metagenomes. FEMS Microbiol Ecol 92:1–8. https://doi.org/10.1093/femsec/fiv168.

43. Sturgill MG, Rapp RP. 1992. Clarithromycin: review of a new macrolide

antibiotic with improved microbiologic spectrum and favorable pharmacokinetic and adverse effect profiles. Ann Pharmacother 26:1099–1108. https://doi.org/10.1177/106002809202600912.

44. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl Environ Microbiol 77:3846–3852. https://doi.org/10.1128/AEM.02772-10.

45. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–1414. https://doi.org/10.1111/1462-2920.13023.

46. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12:38. https://doi.org/10.1186/1471-2105-12-38.

47. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang K, Bin Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9.

48. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

49. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

50. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

51. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: community ecology package. R package version 2.5-6. https://cran.r-project.org/web/packages/vegan/index.html.

52. De Cáceres M, Legendre P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574. https://doi.org/10.1890/08-1823.1.

53. Petrenko P, Lobb B, Kurtz DA, Neufeld JD, Doxey AC. 2015. MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. BMC Biol 13:92. https://doi.org/10.1186/s12915-015-0195-4.

54. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373. https://doi.org/10.1093/nar/gkg128.

55. Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. 2019. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. bioRxiv https://doi.org/10.1186/s12859-020-03585-4.

56. Bushnell B. 2018. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. https://jgi.doe.gov/data-and-tools/bbtools/.

57. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.

58. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

59. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.

60. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

61. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. bioRxiv https://doi.org/10.1101/530188.

62. NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44:D7–D19. https://doi.org/10.1093/nar/gkv1290.

63. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.

64. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.

65. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.

66. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.

67. Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. J Stat Softw 86:1–17. https://doi.org/10.18637/jss.v086.i01.

68. R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

69. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2252. https://doi.org/10.1093/bioinformatics/btz859.

70. Graham ED, Heidelberg JF, Tully BJ. 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. ISME J 12:1861–1866. https://doi.org/10.1038/s41396-018-0091-3.

71. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. 2015. High-specificity targeted functional profiling in microbial communities with ShortBRED. PLoS Comput Biol 11:e1004557. https://doi.org/10.1371/journal.pcbi.1004557.

72. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13.