



Characterization of *Legionella* Species from Watersheds in British Columbia, Canada

Michael A. Peabody,^a Jason A. Caravas,^b Shatavia S. Morrison,^b
Jeffrey W. Mercante,^b Natalie A. Prystajecy,^{c,d} Brian H. Raphael,^b
Fiona S. L. Brinkman^a

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada^a; Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA^b; Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada^c; British Columbia Centre for Disease Control Public Health Laboratory, Vancouver, British Columbia, Canada^d

ABSTRACT *Legionella* spp. present in some human-made water systems can cause Legionnaires' disease in susceptible individuals. Although legionellae have been isolated from the natural environment, variations in the organism's abundance over time and its relationship to aquatic microbiota are poorly understood. Here, we investigated the presence and diversity of legionellae through 16S rRNA gene amplicon and metagenomic sequencing of DNA from isolates collected from seven sites in three watersheds with varied land uses over a period of 1 year. *Legionella* spp. were found in all watersheds and sampling sites, comprising up to 2.1% of the bacterial community composition. The relative abundance of *Legionella* tended to be higher in pristine sites than in sites affected by agricultural activity. The relative abundance levels of *Amoebosoa*, some of which are natural hosts of legionellae, were similarly higher in pristine sites. Compared to other bacterial genera detected, *Legionella* had both the highest richness and highest alpha diversity. Our findings indicate that a highly diverse population of legionellae may be found in a variety of natural aquatic sources. Further characterization of these diverse natural populations of *Legionella* will help inform prevention and control efforts aimed at reducing the risk of *Legionella* colonization of built environments, which could ultimately decrease the risk of human disease.

IMPORTANCE Many species of *Legionella* can cause Legionnaires' disease, a significant cause of bacterial pneumonia. *Legionella* in human-made water systems such as cooling towers and building plumbing systems are the primary sources of Legionnaires' disease outbreaks. In this temporal study of natural aquatic environments, *Legionella* relative abundance was shown to vary in watersheds associated with different land uses. Analysis of the *Legionella* sequences detected at these sites revealed highly diverse populations that included potentially novel *Legionella* species. These findings have important implications for understanding the ecology of *Legionella* and control measures for this pathogen that are aimed at reducing human disease.

KEYWORDS Legionella, metagenomics, watersheds

Legionnaires' disease (LD) is a potentially fatal form of bacterial pneumonia caused by various species of *Legionella* (1). Individuals with chronic lung diseases, immune system deficiencies, smokers, and people of advanced age are at an increased risk for LD. A milder form of legionellosis characterized by fever and "flu-like" symptoms is termed Pontiac fever. In built environments, *Legionella* can multiply in water that is stagnant, maintained at permissive temperatures (~25 to 37°C), and that lacks appropriate disinfectant levels. Various devices, such as cooling towers, showerheads, foun-

Received 29 May 2017 Accepted 11 July 2017 Published 2 August 2017

Citation Peabody MA, Caravas JA, Morrison SS, Mercante JW, Prystajecy NA, Raphael BH, Brinkman FSL. 2017. Characterization of *Legionella* species from watersheds in British Columbia, Canada. *mSphere* 2:e00246-17. <https://doi.org/10.1128/mSphere.00246-17>.

Editor Garret Suen, University of Wisconsin—Madison

Copyright © 2017 Peabody et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Fiona S. L. Brinkman, brinkman@sfu.ca.

tains, and spas, can aerosolize contaminated water. Inhalation of these aerosols by susceptible individuals can result in legionellosis.

Over 60 species of *Legionella* have been identified (<http://www.bacterio.net/-allnamesdl.html>) and at least one-third of these have been linked to human disease (2). *L. pneumophila* is the most frequent cause of LD in North America. Other less common, clinically relevant species include *L. longbeachae*, *L. bozemanii*, *L. micdadei*, and *L. dumoffii* (3–6). *L. pneumophila* is highly diverse, with 17 known serogroups (1). Genome sequence analysis has revealed that much of the genetic diversity among isolates of *L. pneumophila* is driven by recombination (7).

During legionellosis outbreak investigations, *Legionella* isolates from potential environmental sources are compared with clinical isolates in an effort to support epidemiological associations. Various subtyping schemes have been used for this purpose, such as pulsed-field gel electrophoresis, sequence-based typing, and more recently, whole-genome sequencing (1, 8). Confirmation of the environmental sources of *Legionella* may help to shorten the duration of an outbreak by focusing remediation efforts on a specific source and by informing ongoing prevention strategies.

The mechanisms by which sources of *Legionella* from the natural environment colonize the built environment are poorly understood. However, it is likely that at least some of the *Legionella* strains present in source water for built environments is derived from natural aquatic ecosystems, such as rivers, streams, and lakes, where *Legionella* strains have been shown to be widely distributed (9, 10). Various studies have demonstrated a link between *Legionella* and various protozoa, including amoebae such as *Acanthamoeba* spp., *Naegleria* spp., and *Hartmannella* spp. (11, 12). Many of the molecular mechanisms that legionellae use for growth in amoebae appear to overlap those for growth in human macrophages (13). Moreover, growth within amoebae not only amplifies the number of *Legionella* organisms but also may enhance bacterial virulence (13, 14). Finally, *Legionella* spp. have been detected in biofilms, which are considered a major reservoir of the organism in colonized human-made water systems (15, 16).

Few studies have attempted a comprehensive analysis of the microbiome of natural aquatic environments. Recently, a year-long study was conducted to understand the microbial community composition in various watersheds with different land uses in British Columbia, Canada (17, 18). In this study, sites within agricultural, urban, and protected watersheds were sampled monthly. Size fractionation methods were employed to generate templates for sequencing. 16S and 18S amplicon sequencing was conducted to quantify changes in the microbiomes of these environments, while shotgun metagenomic sequencing was conducted to understand community structure and function. Notably, this study revealed that the most abundant bacterial phyla present in the watersheds studied were *Proteobacteria* (which includes *Legionella*), *Actinobacteria*, *Firmicutes*, and *Bacteroidetes* (17).

In order to better understand the ecology of *Legionella* in the natural aquatic environment, we evaluated this extensive data set with the goals of quantifying *Legionella* abundance among different watersheds, determining the diversity of *Legionella* spp. present, and evaluating the role of amoeba when in the presence of this bacterium in such environments. Understanding the presence and diversity of *Legionella* in these watersheds may help to improve our ability to control colonization by these organisms of human-made water systems from natural water sources.

RESULTS

Detection of *Legionella* over time at various sampling sites. Analysis of the 16S data set revealed that *Legionella* spp. were found in all watersheds and sampling sites (Fig. 1). Overall, the relative abundances of *Legionella* spp. were $\leq 2.1\%$ of the bacterial taxa present and were significantly higher (q -value, < 0.05) (see Table S1 in our supplementary files posted on figshare [<https://doi.org/10.6084/m9.figshare.5046937>]) in samples from more-pristine sites (protected upstream [PUP] samples and upstream of agricultural activity [AUP] samples) than from sites affected by agricultural activity,

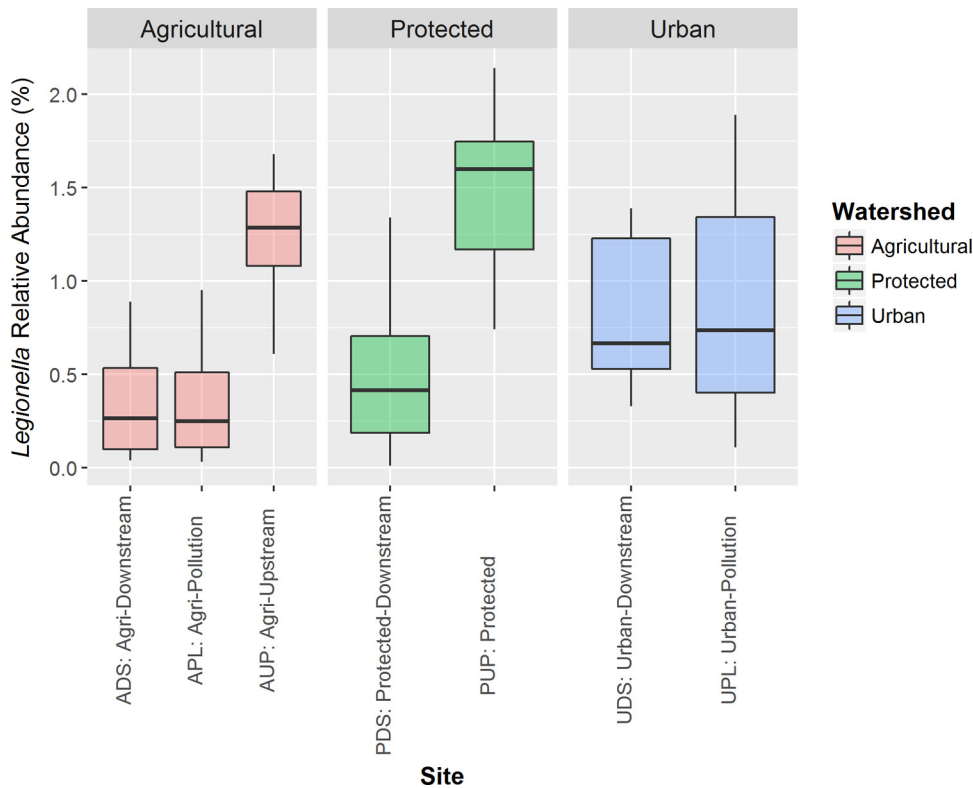


FIG 1 Abundance of *Legionella* species at various sites. The box plots represent the relative abundances of *Legionella* present for all samples obtained from the indicated site, which was determined by 16S rRNA sequencing. Boxes are bounded by 25th and 75th percentiles, and the middle line represents the median relative abundance.

while the sites affected by urban activity had intermediate relative abundance levels of *Legionella* spp. The relative abundance of *Legionella* spp. varied substantially over the sampling period (see Fig. S1).

Distribution of *Legionella* species among sampling sites. To examine the distribution of *Legionella* spp. at a higher resolution than the genus level, the shotgun metagenomics data set was used to obtain species-level classification. The shotgun metagenomics reads were classified to >40 *Legionella* species by using MEGAN6, albeit with very few reads associated with some species (Fig. 2). A majority of sequence reads were assigned only at the *Legionella* genus level. Among those reads identified at the species level, *L. pneumophila* was the most abundant at all sampling sites.

In addition, we used a BLAST approach to identify sequence reads from the shotgun metagenomics data set matching the *mip* gene, which is used extensively for *Legionella* species determinations (19). This analysis also found a low number of reads matching >35 different *Legionella* spp. among the samples examined (see Table S2). Notably, the range of nucleotide identity among the reads aligned with known *mip* genes varied widely (~68 to 94%).

Diversity of *Legionella* isolates relative to other bacterial taxa. To better understand the diversity of *Legionella* spp. present in the samples, the 16S rRNA gene data sets were examined at the operational taxonomic unit (OTU) level. OTUs approximate microbial taxa, but are not limited to previously sequenced taxa (as in the MEGAN6 analysis), thus providing a useful complement to the species-level classification performed with metagenomics data. A total of 71 OTUs were assigned to *Legionella*, supporting the view of a highly diverse population of *Legionella* spp. present in these samples (Fig. 3A). *Legionella* is the ninth most abundant genus in the data set, so to rule out the possibility of the richness being driven by relative abundance, the richness of *Legionella* was compared to other abundant genera. The top 50 most abundant genera in the data set over all sampling sites, based on the sum of all reads assigned to the

Reads assigned to *Legionella* with MEGAN6

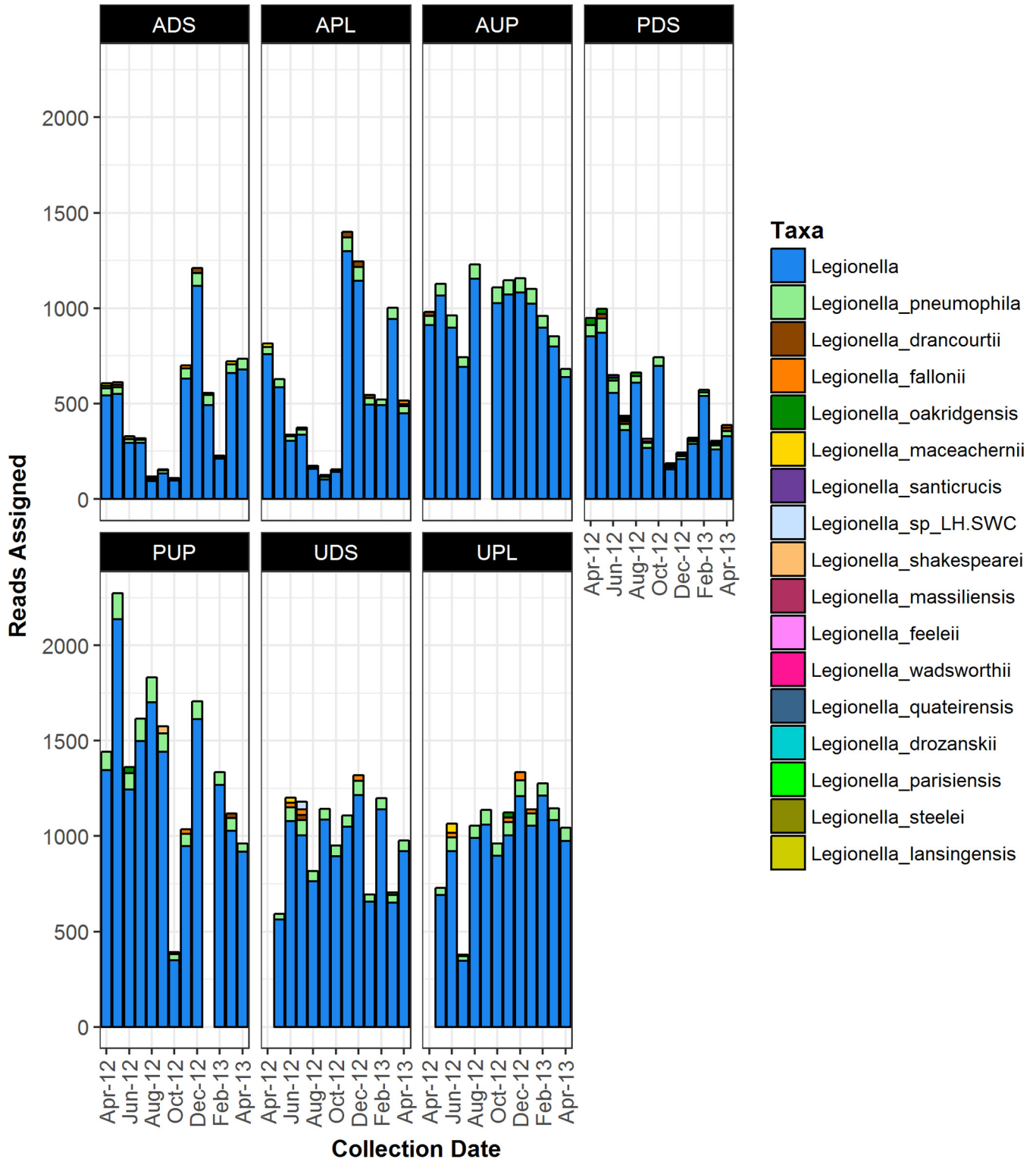


FIG 2 Taxonomic classification of *Legionella* metagenome shotgun sequencing reads. The numbers of metagenomic sequencing reads classified by MEGAN6 analysis to *Legionella* are shown for each sampling site and date of collection. Reads assigned to *Legionella* spp. for which the species-level assignment comprised fewer than 2% of the total reads assigned to *Legionella* are denoted “Legionella” in the legend (i.e., “Legionella” denotes the sum of all reads assigned to the genus *Legionella* and all *Legionella* species not depicted separately). Species in the legend are ordered from most abundant to least abundant in the overall data set (all samples).

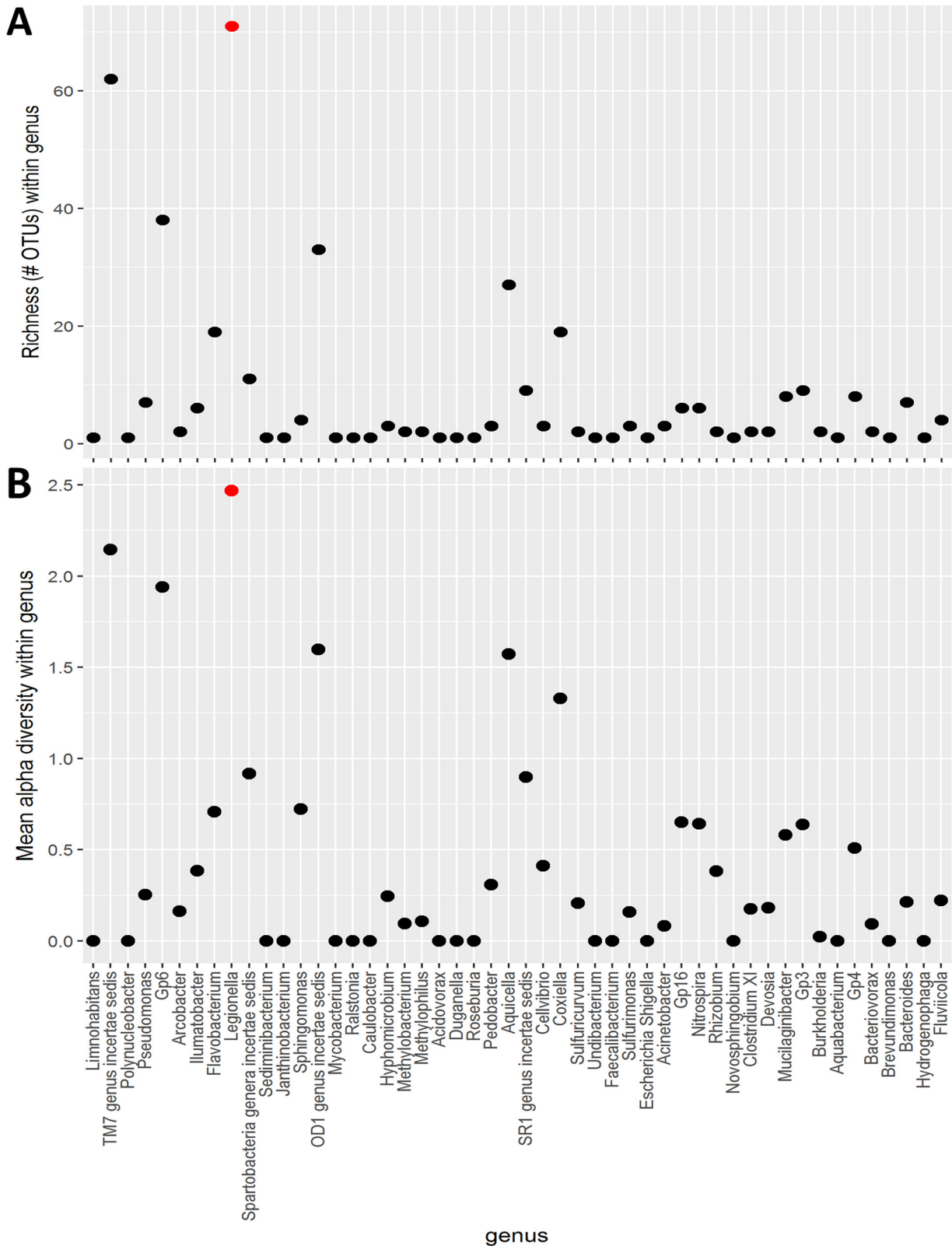


FIG 3 Comparison of diversity levels of selected bacterial taxa. The top 50 most abundant genera derived from 16S rRNA sequencing are shown in order of greatest to least abundance, from left to right. (A) The number of OTUs (richness) of each taxon is shown. (B) Mean alpha diversity was calculated based on the OTUs assigned to each genus shown. Richness and alpha diversity values for *Legionella* are shown by the red dots in each chart.

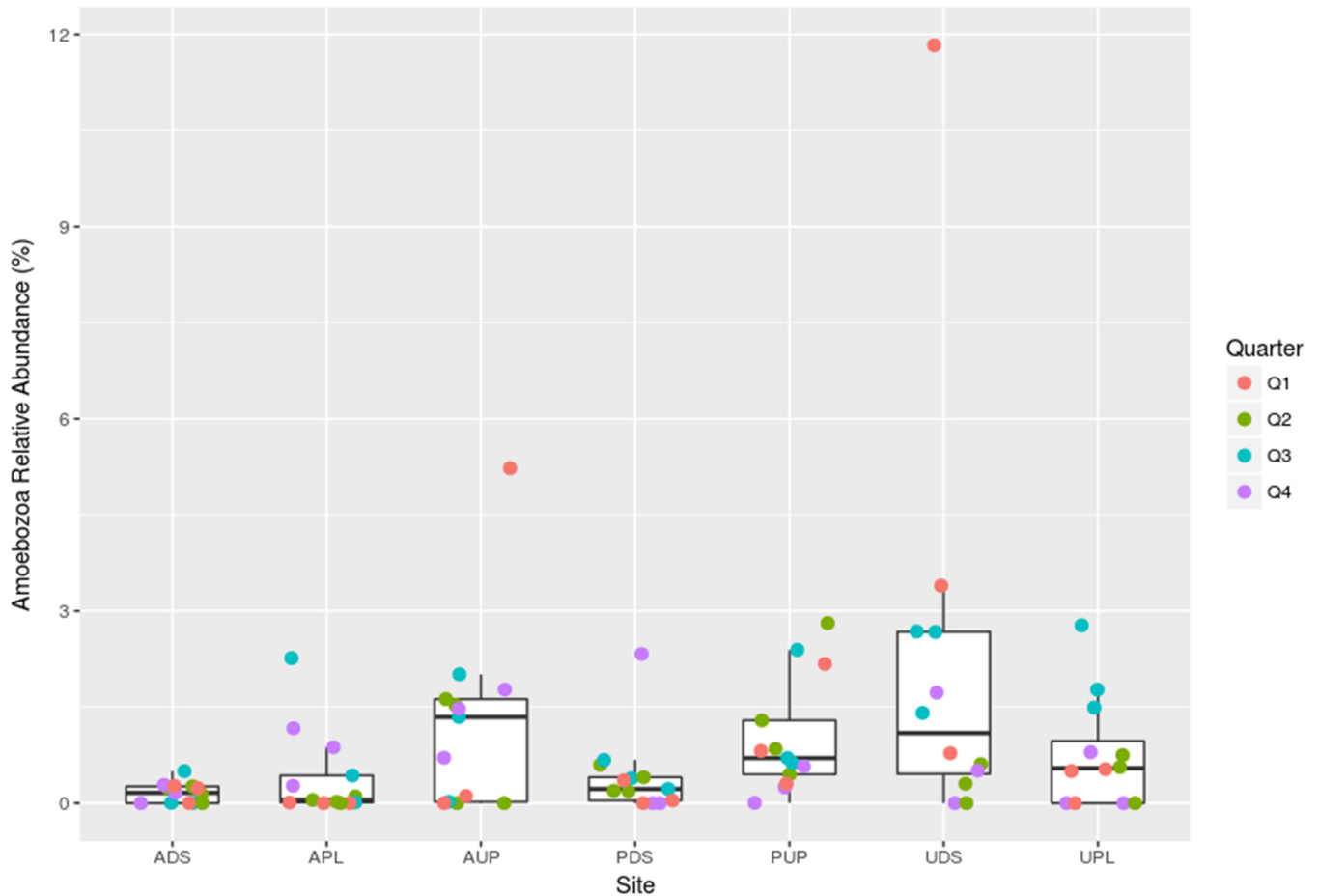


FIG 4 Relative abundance of *Amoebzoza* at various sites. Box plots represent the relative abundances of *Amoebzoza* present for all samples obtained from the indicated site, determined by 18S rRNA sequencing. Each circle represents the relative abundance of *Amoebzoza* in a specific sample. Circles are color coded by the date of sample collection, where quarters of the year are designated as follows: Q1, January, February, and March; Q2, April, May, and June; Q3, July, August, and September; Q4, October, November, and December. Boxes are bounded by the 25th and 75th percentiles, and the middle line represents the median relative abundance.

genera from all samples, were plotted versus the richness or number of OTUs assigned to that genus (see Fig. S2 in our figshare files). Richness was unevenly distributed among the taxa and did not seem to be driven by relative abundance. Most genera had a richness of fewer than 10 OTUs, and only 5 genera had a richness greater than 20 OTUs.

Legionella had markedly higher richness than the mean (8.1 OTUs) and median (2 OTUs) richness observed among the top 50 most abundant genera. This high richness of *Legionella* was even more pronounced compared to the entire data set (versus just the top 50), which had a mean richness of 3.2 and a median of 1. The richness of OTUs assigned to *Legionella* was distributed across sample sites and time (see Fig. S3). Finally, average alpha diversity values over all samples of the 50 most abundant genera were also plotted and showed that *Legionella* was not only the richest but also the most diverse bacterial genus identified (Fig. 3B).

Comparison of *Legionella* and *Amoebzoza* relative abundances. Analysis of the 18S rRNA gene data set revealed a wide distribution in the relative abundance levels of the phylum *Amoebzoza* among the sample sites (Fig. 4). A significant difference (q -value, <0.05) was observed between the site downstream of agricultural activity (ADS) and the PUP and urban downstream (UDS) sites (see Table S3). The more pristine sites within the AUP and PUP watersheds displayed wider distributions of *Amoebzoza* relative abundance levels than the more impacted sites (ADS and PDS [the site downstream from the protected reservoir]). The most pronounced seasonal effect was

observed in the urban floodplain (UPL) site, where samples collected between July and September had the highest relative abundance of *Amoebozoa*. (See Materials and Methods for further descriptions of the sampling sites.)

Platyamoeba was the most abundant genus (mean relative abundance over all samples, ~0.3%) represented among the *Amoebozoa* (data not shown). *Acanthamoeba*, a well-known host of *Legionella*, had a mean relative abundance that was nearly 100-fold lower than that of *Platyamoeba*. *Naegleria*, an amoebal member of a separate phylum which can also host *Legionella*, was not detected in any samples.

DISCUSSION

The robust methodology described by Uyaguari-Diaz et al. (17) to separate various microbial components (eukaryotic, bacterial, and viral) in natural water samples via both amplicon (16S and 18S) as well as metagenomic sequencing was used to characterize the compositions of water samples from various watersheds in British Columbia. The predominant bacterial phyla in the analysis of seven watershed sampling sites using metagenomic sequencing were *Proteobacteria*, *Actinobacteria*, *Firmicutes*, and *Bacteroidetes*. A year-long study that examined the same sites showed a shift in the microbial community composition (revealed by the average genome composition and k-mer composition) among some of the sites corresponding to the season and/or nutrient concentrations (18). In the current study, we detected *Legionella*, a member of the *Gammaproteobacteria* order, at all sampling sites and collection dates.

Unlike the seasonal patterns observed when examining the composition of the entire bacterial community (18), the relative abundance of *Legionella* varied throughout the year, without any discernible seasonal patterns, reaching as high as 2% of the bacterial taxa present. Although LD cases peak in summer and autumn months, natural waters are not considered a source of disease to the extent that water systems in built environments are considered the sources (1). Other researchers have detected *Legionella* by using culture or quantitative PCR in natural water sources. In a recent study, investigators found ~10⁴ to 10⁵ cells/liter of *Legionella* spp. in some Taiwanese river water samples via real-time PCR (20), while a previous study of marine and freshwater sites in Puerto Rico demonstrated an abundance of *L. pneumophila* of 10⁴ cells/ml via direct fluorescence antibody (DFA) testing (21). These findings suggest that the quantity of *Legionella* in the natural environment may be highly variable.

The relative abundance of *Legionella* was highest in sites with limited land use: from the site upstream of agricultural activity (AUP) and from a river that empties into a drinking water reservoir (PUP). In both cases, these sources feed downstream sites where the relative abundance of *Legionella* is lower. There are several possible explanations for this decrease in *Legionella* relative abundance. Downstream sites may contain contaminants or lack specific nutrients for *Legionella* growth. There may also be an increase in certain non-*Legionella* genera in these downstream sites, resulting in a lower relative abundance of *Legionella* in the community. Alternatively, *Legionella* may become associated with biofilms, which would decrease their relative abundance in the surface water samples collected in this study. Notably, the water collected at the PDS site travels through a nearly 9-km pipe made of concrete and steel and lined with coal tar. It is possible that biofilms present within this pipe may trap *Legionella*; *Legionella* can also survive and grow within various amoeba species. Similar to the pattern observed with *Legionella* relative abundances, the lowest mean relative abundances of *Amoebozoa* were found at the ADS, APL, and PDS sites, supporting the possibility that *Legionella* relative abundance is amplified by the presence of amoeba in these natural water sources. The differences in relative abundance of *Legionella* seen in our study may also be due to the presence of other protozoa; in addition to amoebas, *L. pneumophila* has been shown to infect and grow within ciliates (22), and protozoan predators have recently been isolated that graze on virulent *Legionella* spp. (23).

Note that the abundance values for both bacteria and eukaryotes were relative rather than absolute abundance values. Furthermore, the relative abundance of eukaryotes inferred by 18S rRNA gene sequencing would be affected by the large

variation in copy numbers of the 18S rRNA gene, which can vary by several orders of magnitude between species (24).

This study revealed that *Legionella* spp. present in the watersheds examined are incredibly diverse. More than 70 OTUs were detected via 16S amplicon sequencing. The metagenomic sequence analysis used in this study demonstrated that *L. pneumophila* was the most common species represented. Similarly, Fliermans et al. detected *L. pneumophila* by DFA in nearly all concentrated water samples collected from 67 natural water sources in North Carolina, South Carolina, Georgia, Florida, Alabama, Indiana, and Illinois (10). Various *Legionella* spp. are frequently detected in studies of natural water sources (20, 21, 25, 26). Notably, sequence analysis of the most common *Legionella* 16S rRNA gene-based OTU amplified from water samples along a French river were associated with unknown/uncultured bacteria (25). The high diversity of *Legionella* spp. among these sources may have implications for clinical disease, since several non-*pneumophila* *Legionella* species are associated with clinical disease (including pneumonia), especially among immunosuppressed populations (2). A study of natural water sources in the Mount St. Helens (Washington, USA) blast zone was conducted after researchers exposed to lakes and streams in the region reported symptoms consistent with Pontiac fever in the early 1980s (27). Various known *Legionella* spp. were detected in this study, with higher organism relative abundances found in water samples taken within the blast zone and in lakes receiving water from hydrothermal seeps than in sites outside the blast zone. A novel species (*L. sainthelensi*) was isolated from water samples collected around Mount St. Helens (28), and this species was subsequently found to be associated with clinical disease (29). Although we did not attempt to isolate and grow the putative novel *Legionella* species from our samples, doing so could be the next step for future studies.

Metagenomics classification programs such as MEGAN6 used in this study may overclassify reads to incorrect species if the matching species is not present in the database (30). More specifically, the program might assign reads to the most closely related species in the database. There are currently over 500 *L. pneumophila* genome sequences in the NCBI database, but only a few representatives are present for other *Legionella* species. The wide range of alignment identities observed with the MEGAN6 analysis further suggests that novel or uncharacterized *Legionella* strains may be present in the samples. Notably, alignment of the shotgun metagenomic reads with the *Legionella mip* gene also uncovered a large number of *Legionella* species (>35) among the watershed samples, but sequencing coverage of this gene may be limited. Nonetheless, the alignment identity of these matches was low (typically <80%), further suggesting the presence of additional novel *Legionella* species in these watersheds.

While the presence of *Legionella* spp. in natural water samples alone is not a significant public health concern, these organisms may seed human-made water systems. In turn, these systems could become sources of *Legionella* dissemination under permissive conditions. Understanding the diversity of organisms present in the natural aquatic environment and factors that may contribute to increased abundance of specific *Legionella* spp. in these environments may help public health workers identify potential new threats to human health and respond quickly to LD by using improved diagnostic and typing assays. This study demonstrates that natural aquatic environments, including watersheds, likely harbor previously unrecognized *Legionella* spp. As culture-independent diagnostic tests for LD become more commonly utilized, it will be important to evaluate the ability of these assays to detect new and emerging *Legionella* spp. and assess their potential to cause disease.

MATERIALS AND METHODS

Environmental sampling sites and processing. Water sample collection and processing have been described previously (17). Water samples were collected monthly from seven sites in three watersheds in southwestern British Columbia, Canada. Contextual data for the samples can be found on the figshare website (https://figshare.com/articles/Sample_data_MlxS_format/5188063). The watersheds had varied land uses: an agricultural watershed, an urban watershed, and a protected watershed that was used as a drinking water source. Three sites within the agricultural watershed were sampled, including a site

upstream of agricultural activity (AUP), within a highly farmed and irrigated floodplain (APL), and a site downstream of this activity (ADS). Two sites were sampled in the protected and urban watersheds. A forested and protected river site that empties into a drinking water reservoir (PUP) was sampled, along with a site where water from the reservoir empties out of a 9-km-long pipe (PDS). Finally, urban sites selected from a stream passing through 300 m (UPL) and 1 km downstream (UDS) in a residential development were sampled. Samples were collected between April 2012 and April 2013 for the protected (PUP and PDS) and agriculturally affected (AUP, APL, and ADS) samples. The urban affected samples (UPL and UDS) were collected from May 2012 to April 2013.

For each sample, 40 liters of water was collected and filtered in the field through a 105- μm -pore-size Spectra/mesh polypropylene filter to remove larger particles and debris. Samples were kept on ice and transported to the British Columbia Centre for Disease Control Public Health Laboratory (BCCDC PHL) and were processed within 24 h. Samples were serially filtered to generate size-specific fractions relating to microbial class. A 1- μm -pore-size filter (Envirochek HV; Pall Corporation, Ann Arbor, MI) was used to capture eukaryote-sized particles, followed by a 0.2- μm , 142-mm Supor-200 membrane disk filter (Pall Corporation, Ann Arbor, MI) to capture bacterial and archaeal-sized particles. Eukaryotic-sized cells retained in the 1- μm Envirochek HV capsules were eluted per the manufacturer's protocol. For this fraction, before extraction and to facilitate disruption of eukaryotic cells, eight freeze-thaw cycles were followed by digestion overnight with proteinase K (Qiagen Sciences, Germantown, MD). DNA from both the eukaryote- and bacteria-sized fractions were extracted using the PowerLyzer Powersoil DNA isolation kit (MoBio, Inc., Carlsbad, CA), which uses a combination of bead beating and chemical lysis. Particle-associated bacterial cells would have been removed by the 105- μm and 1- μm filters, and so the microbial community collected from the 0.2- μm filter was primarily composed of the free-living bacterioplankton community. This filtering thus misses microorganisms that are particle associated or are filtered out due to being larger than the filter sizes used.

Shotgun and amplicon sequencing. Both shotgun and amplicon sequencing were performed with the Illumina MiSeq (Illumina, Inc., San Diego, CA) using the MiSeq reagent kit V2 (two 250-bp paired-end reads, 500 cycles) at the BCCDC PHL. Each sequence run included a positive control (mock community) and negative control (ultrapure water) sample, as described previously (17). Amplicons were generated as previously described, targeting the V1 to V3 regions of the 18S rRNA gene for the eukaryote-sized fraction and targeting V3 and V4 regions of the 16S rRNA gene for the bacteria-sized fraction (17). Amplicons were purified using the QIAquick PCR purification kit (Qiagen Sciences, Maryland, MD), and sequencing libraries were prepared using the NEXTflex ChIP-Seq kit (BIOO Scientific, Austin, TX) using the gel size selection option, both per the manufacturer's instructions. The shotgun sequencing libraries were prepared using the Nextera XT DNA sample preparation kit (Illumina, Inc., San Diego, CA), and gel-size selection was performed using the Ranger technology (Coastal Genomics Inc., Burnaby, BC, Canada) and targeting fragments between 500 and 800 bp (31).

Bioinformatic analysis. Quality control for the 16S sequences was performed using Cutadapt (32) and Trimmomatic (33) with default parameters and removing reads from the data set if, following quality filtering and trimming, either read in the pair was less than 200 bp. Following this processing, samples were subsampled to 10,000 reads. Samples that had fewer than 10,000 reads were removed from further analysis (the omitted samples included the May 2012 samples from the ADS, PDS, and UDS sites, the September 2012 sample from the PUP site, and the December 2012 sample from the UDS site). The MiSeq protocol in Mothur was used to generate OTUs and their taxonomic assignments (34, 35). The data were exported in BIOM format (36), and the OTUs were extracted that were assigned to *Legionella*.

Shotgun sequencing reads were first trimmed to remove low-quality bases by using Trimmomatic (33). Trimming at the 3' end was performed with a sliding window of 5 bases and a minimum Phred score of 20, and trimming at the 5' end was performed with consecutive bases with Phred scores of less than 20 removed. Sequencing adapters were then removed using CutAdapt (32), overlapping paired-end reads were merged with PEAR (37), and reads shorter than 100 bp were discarded. To normalize the samples, all samples were subsampled to 418,500 reads, the smallest number of reads in a sample (sample 32). The January 2013 sample from the PUP site was omitted from analyses due to too few reads (sample 82). Sequence reads were then run on Diamond version 0.8.36 (38) with default parameters against the NCBI nr database (downloaded 21 February 2017). These results were used as input for taxonomic classification by MN version 6.6.7 (39) using default parameters, except that the Min Support percentage which was set to 0. The taxonomic classification files were then parsed to extract the assignments directly to *Legionella* at the genus level or to any specific *Legionella* species. Additionally, a custom BLAST database containing >700 *Legionella* spp. *mip* nucleotide sequences was used to identify raw reads aligning to at least 100 nucleotides and an *E* value of 1×10^{-10} with this gene. BLAST using the NCBI nt database was also performed with the identified reads, and those with better alignments to non-*Legionella* targets were excluded.

Analyses were performed in R (v3.2.3), and Shannon's diversity index was calculated using the vegan package (40). The Kruskal-Wallis method with Dunn's test was used to compare the relative abundance of legionellae among the sites, with *P* values adjusted using the Benjamini-Hochberg method. The September 2012 sample from the AUP site was removed from analyses, as it had been previously noted as unusual and perhaps mislabeled during sample processing (18).

Paired-end 18S read files were quality filtered and trimmed with Trim Galore v0.3.7 (available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) by using a quality score cutoff of 25. Remaining read pairs with more than 10 overlapping nucleotides were joined using FLASH v1.2.11 (41). The merged read pairs were used as input to QIIME v1.9.1 (42), with the SILVA (release 128) 18S database (43) used for chimera filtering and open reference OTU assignment.

Accession number(s). Sequence data were submitted to the NCBI Sequence Read Archive under BioProject ID [287840](https://www.ncbi.nlm.nih.gov/bioproject/287840).

ACKNOWLEDGMENTS

We thank Claressa Lucas for critical review of the manuscript.

Funding for this work was primarily provided by Genome Canada and Genome BC (LSARP-165WAT). M.A.P. was funded by a National Sciences and Engineering Research Council of Canada Postgraduate Scholarship and a Canadian Institute of Health Research Strategic Training Program in Bioinformatics Graduate Award.

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

- Mercante JW, Winchell JM. 2015. Current and emerging *Legionella* diagnostics for laboratory and outbreak investigations. *Clin Microbiol Rev* 28:95–133. <https://doi.org/10.1128/CMR.00029-14>.
- Muder RR, Yu VL. 2002. Infection due to *Legionella* species other than *L. pneumophila*. *Clin Infect Dis* 35:990–998. <https://doi.org/10.1086/342884>.
- Amodeo MR, Murdoch DR, Pithie AD. 2010. Legionnaires' disease caused by *Legionella longbeachae* and *Legionella pneumophila*: comparison of clinical features, host-related risk factors, and outcomes. *Clin Microbiol Infect* 16:1405–1407. <https://doi.org/10.1111/j.1469-0691.2009.03125.x>.
- Fang GD, Yu VL, Vickers RM. 1989. Disease due to the Legionellaceae (other than *Legionella pneumophila*). Historical, microbiological, clinical, and epidemiological review. *Medicine* 68:116–132. <https://doi.org/10.1097/00005792-198903000-00005>.
- Reingold AL, Thomason BM, Brake BJ, Thacker L, Wilkinson HW, Kuritsky JN. 1984. *Legionella pneumonia* in the United States: the distribution of serogroups and species causing human illness. *J Infect Dis* 149:819. <https://doi.org/10.1093/infdis/149.5.819>.
- Yu VL, Plouffe JF, Pastoris MC, Stout JE, Schousboe M, Widmer A, Summersgill J, File T, Heath CM, Paterson DL, Chereschsky A. 2002. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *J Infect Dis* 186:127–128. <https://doi.org/10.1086/341087>.
- Sánchez-Busó L, Comas I, Jorques G, González-Candelas F. 2014. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet* 46:1205–1211. <https://doi.org/10.1038/ng.3114>.
- David S, Mentasti M, Tewolde R, Aslett M, Harris SR, Afshar B, Underwood A, Fry NK, Parkhill J, Harrison TG. 2016. Evaluation of an optimal epidemiological typing scheme for *Legionella pneumophila* with whole-genome sequence data using validation guidelines. *J Clin Microbiol* 54:2135–2148. <https://doi.org/10.1128/JCM.00432-16>.
- Morris GK, Patton CM, Feeley JC, Johnson SE, Gorman G, Martin WT, Skaliy P, Mallison GF, Politi BD, Mackel DC. 1979. Isolation of the Legionnaires' disease bacterium from environmental samples. *Ann Intern Med* 90:664–666. <https://doi.org/10.7326/0003-4819-90-4-664>.
- Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. 1981. Ecological distribution of *Legionella pneumophila*. *Appl Environ Microbiol* 41:9–16.
- Rowbotham TJ. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Pathol* 33:1179–1183. <https://doi.org/10.1136/jcp.33.12.1179>.
- Fields BS. 1993. *Legionella* and protozoa: interaction of a pathogen and its natural host, p 129–136. In Barbaree JM, Breiman RF, Dufour AP (ed), *Legionella: current status and emerging perspectives*. American Society for Microbiology, Washington, DC.
- Swanson MS, Hammer BK. 2000. *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. *Annu Rev Microbiol* 54:567–613. <https://doi.org/10.1146/annurev.micro.54.1.567>.
- Cirillo JD, Falkow S, Tompkins LS. 1994. Growth of *Legionella pneumophila* in *Acanthamoeba castellanii* enhances invasion. *Infect Immun* 62:3254–3261.
- Declerck P. 2010. Biofilms: the environmental playground of *Legionella pneumophila*. *Environ Microbiol* 12:557–566. <https://doi.org/10.1111/j.1462-2920.2009.02025.x>.
- Taylor M, Ross K, Bentham R. 2009. *Legionella*, protozoa, and biofilms: interactions within complex microbial systems. *Microb Ecol* 58:538–547. <https://doi.org/10.1007/s00248-009-9514-z>.
- Uyaguari-Diaz MI, Chan M, Chaban BL, Croxson MA, Finke JF, Hill JE, Peabody MA, Van Rossum T, Suttle CA, Brinkman FSL, Isaac-Renton J, Prystajek NA, Tang P. 2016. A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples. *Microbiome* 4:20. <https://doi.org/10.1186/s40168-016-0166-1>.
- Van Rossum T, Peabody MA, Uyaguari-Diaz MI, Cronin KI, Chan M, Slobodan JR, Nesbitt MJ, Suttle CA, Hsiao WWL, Tang PKC, Prystajek NA, Brinkman FSL. 2015. Year-long metagenomic study of river microbiomes across land use and water quality. *Front Microbiol* 6:1405. <https://doi.org/10.3389/fmicb.2015.01405>.
- Ratcliff RM, Lanser JA, Manning PA, Heuzenroeder MW. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the mip gene. *J Clin Microbiol* 36:1560–1567.
- Kao PM, Tung MC, Hsu BM, Chiu YC, She CY, Shen SM, Huang YL, Huang WC. 2013. Identification and quantitative detection of *Legionella* spp. in various aquatic environments by real-time PCR assay. *Environ Sci Pollut Res Int* 20:6128–6137. <https://doi.org/10.1007/s11356-013-1534-z>.
- Ortiz-Roque CM, Hazen TC. 1987. Abundance and distribution of Legionellaceae in Puerto Rican waters. *Appl Environ Microbiol* 53:2231–2236.
- Fields BS, Shotts EB, Feeley JC, Gorman GW, Martin WT. 1984. Proliferation of *Legionella pneumophila* as an intracellular parasite of the ciliated protozoan *Tetrahymena pyriformis*. *Appl Environ Microbiol* 47:467–471.
- Amaro F, Wang W, Gilbert JA, Anderson OR, Shuman HA. 2015. Diverse protist grazers select for virulence-related traits in *Legionella*. *ISME J* 9:1607–1618. <https://doi.org/10.1038/ismej.2014.248>.
- Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46:48–50. <https://doi.org/10.1139/g02-103>.
- Parthuisot N, West NJ, Lebaron P, Baudart J. 2010. High diversity and abundance of *Legionella* spp. in a pristine river and impact of seasonal and anthropogenic effects. *Appl Environ Microbiol* 76:8201–8210. <https://doi.org/10.1128/AEM.00188-10>.
- Carvalho FRS, Natasí FR, Gamba RC, Foronda AS, Pellizari VH. 2006. Diversity of *Legionella* spp. in Antarctic lakes of the Keller peninsula, p 417–419. In Cianciotto NP, Kwak YA, Edelstein PH, Fields BS (ed), *Legionella: state of the art 30 years after its recognition*. American Society for Microbiology, Washington, DC.
- Tison DL, Baross JA, Seidler RJ. 1983. *Legionella* in aquatic habitats in the Mount Saint Helens blast zone. *Curr Microbiol* 9:345–348. <https://doi.org/10.1007/BF01588832>.
- Campbell J, Bibb WF, Lambert MA, Eng S, Steigerwalt AG, Allard J, Moss CW, Brenner DJ. 1984. *Legionella sainthelensi*: a new species of *Legionella* isolated from water near Mt. St. Helens. *Appl Environ Microbiol* 47:369–373.
- Benson RF, Thacker WL, Fang FC, Kanter B, Mayberry WR, Brenner DJ. 1990. *Legionella sainthelensi* serogroup 2 isolated from patients with pneumonia. *Res Microbiol* 141:453–463. [https://doi.org/10.1016/0923-2508\(90\)90071-W](https://doi.org/10.1016/0923-2508(90)90071-W).
- Peabody MA, Van Rossum TV, Lo R, Brinkman FS. 2015. Evaluation of shotgun metagenomics sequence classification methods using *in silico*

- and *in vitro* simulated communities. BMC Bioinformatics 16:363. <https://doi.org/10.1186/s12859-015-0788-5>.
31. Uyaguari-Diaz MI, Slobodan JR, Nesbitt MJ, Croxen MA, Isaac-Renton J, Prystajek NA, Tang P. 2015. Automated gel size selection to improve the quality of next-generation sequencing libraries prepared from environmental water samples. J Vis Exp 98:e52685. <https://doi.org/10.3791/52685>.
 32. 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>.
 33. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 34. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
 35. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120. <https://doi.org/10.1128/AEM.01043-13>.
 36. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. Gigascience 1:7. <https://doi.org/10.1186/2047-217X-1-7>.
 37. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. <https://doi.org/10.1093/bioinformatics/btt593>.
 38. 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>.
 39. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R. 2016. MegaN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>.
 40. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2017. Vegan: community ecology package. R package version 2.4-3. Comprehensive R Archive Network, R Foundation for Statistical Computing, Vienna, Austria.
 41. Magoč T, Salzberg SL. 2011. FLASH: fast length of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
 42. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenkov T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
 43. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res 42:D643–DD648. <https://doi.org/10.1093/nar/gkt1209>.