npj | clean water

www.nature.com/npjcleanwater

ARTICLE OPEN

Check for updates

Identification of microbes coexisting with *Legionella* spp. in bathwaters

Masaki Okunaga[1,2], Kenta Kushiro[1], Ryohei Horie[2], Akihiro Kondo[2] and Takashi Abe[1]✉

In Japan, bathwaters are a potential source of legionellosis; therefore, the bathwater in public facilities must be regularly tested for *Legionella* spp. Currently, *Legionella* spp. is detected using the culture method, which is time-consuming and has limited accuracy. Moreover, the reproductive environment of *Legionella* spp. in bathwater remains unclear. Here, we investigated the environment in which *Legionella* spp. reproduce by analyzing the bathwater microbiota and its relationship with *Legionella* spp. We identified the microbiota of 112 bathwater samples collected from bathing facilities by sequencing 16S and 18S rRNA genes. Differences in the microbiota were observed between samples that tested positive and negative for *Legionella* spp., according to 16S rRNA sequencing and culture methods. *Methyloversatilis*, *Cupriavidus*, *Phenylobacterium*, *Vermamoeba*, and *Aspidisca* were highly correlated with *Legionella* spp. Our results support the development of strategies against legionellosis and elucidate the relationships between *Legionella* spp. and the coexisting microbiota in various environments.

npj Clean Water (2022)5:65; https://doi.org/10.1038/s41545-022-00212-y

INTRODUCTION

Legionella spp. are aerobic, gram-negative bacteria that belong to the order Gammaproteobacteria. *Legionella* spp. survive and grow in protozoan hosts, such as amoebae, and form biofilms in natural waters (lakes and rivers) and artificial waters (hot springs, bathing facilities, and cooling towers)[1–3]. Legionnaires' disease is a respiratory illness caused by the aspiration of water or inhalation of aerosols containing *Legionella* spp[1]. More than 60 species of *Legionella* spp. have been identified[4], with *L. pneumophila* being the most frequently detected in patients[5]. The risk of developing Legionnaires' disease is high in immunocompromised and elderly individuals as well as those who smoke[2,6], and the number of affected patients is increasing every year globally. For example, at least 8000 to 18,000 individuals are infected each year in the United States[1], and more than 2000 cases have been reported annually since 2018 in Japan[2]. Bathwaters are a potential source of legionellosis in Japan. In the 2001–2007 survey of legionellosis in Japan reported by Kuroki et al., most patients were elderly individuals (the mean age was 67 years), and more than 40% of infection sources were in public baths[7]. The public bath for the elderly individuals is popular in Japan because of its therapeutic significance. Thus, there are approximately 25,000 public baths in Japan, as reported by the Japanese Ministry of Health, Labor and Welfare (MHLW), and the majority of these use a circulating water system. Young, healthy people with strong immune systems are not susceptible to infection when exposed to *Legionella* spp. Those at a higher risk of infection are elderly individuals, those with underlying diseases, and immunocompromised individuals. Thus, there is a need to establish strict regulations to prevent bacterial contamination in all public baths[2]. The MHLW is committed to ensuring that public baths provide a safe environment and that effective infection control measures are implemented on a regular basis. Strict regulations are in place and enforced to ensure that public baths provide a safe environment and that effective infection control measures are implemented on a regular basis. Thus, Legionnaires' disease remains an important public health threat, especially in the rapidly aging population[2]. Consequently, periodic testing of *Legionella* spp. is mandatory for bathing facilities in Japan.

Legionella spp. can be detected in water samples using culture or molecular methods, such as real-time PCR[8]. The culture method is commonly used[9]; however, it is time-consuming owing to the long incubation period required and can only quantify viable and culturable *Legionella* spp[10]. In addition, the culture method has reliability limitations because its recovery rate is frequently < 100% owing to bacterial loss during the enrichment process[11,12]. The real-time PCR method can provide results faster than the culture method; however, the results of the two methods correlate poorly[10]. The PCR method is considered to overestimate the presence of *Legionella* spp., as it can detect the DNA extracted from dead, viable, and non-culturable *Legionella* cells, including those hosted in amoebae[13,14].

Previous studies have identified several protozoan species in which *Legionella* spp. are capable of intracellular growth, and the presence of these organisms is necessary for the survival and growth of *Legionella* spp. Therefore, it is important to examine the microbiota of the environment in which *Legionella* spp. survive to effectively control Legionnaires' disease; however, the relationship between the survival and growth of *Legionella* spp. and the coexisting microbiota is not fully understood[1]. Peabody et al.[15] and Llewellyn et al.[16] reported metagenomic data of microorganisms containing *Legionella* spp. in environmental water and cooling towers, respectively. These authors showed that the analysis of the flora in these systems is important. On the contrary, there is a lack of extensive metagenomic analysis of microorganisms including *Legionella* spp. in bathwater.

In this study, we investigated the environment in which *Legionella* spp. reproduce by analyzing the bathwater microbiota and its relationship with *Legionella* spp. We investigated the bathwater microbiota using 16S and 18S rRNA sequencing in 112 samples from bathing facilities. We compared the detection results of 16S and 18S rRNA sequencing and culture tests.

[1]Department of Information Engineering, Graduate School of Science and Technology, Niigata University, Niigata-ken 950-2181, Japan. [2]Hiyoshi Corporation, Shiga-ken 523-8555, Japan. ✉email: takaabe@ie.niigata-u.ac.jp

Table 1. Top 10 frequencies of the phylogenetic groups identified by 16S rRNA sequences.

16S rRNA						
	Phylum		Class		Genus	
Ranking	Name	Frequency	Name	Frequency	Name	Frequency
1	Proteobacteria	0.8936	Alphaproteobacteria	0.6101	Methylobacterium	0.1516
2	Bacteroidetes	0.0634	Gammaproteobacteria	0.2794	Sphingomonas	0.1235
3	Actinobacteria	0.0178	Bacteroidia	0.0633	Acinetobacter	0.0783
4	Planctomycetes	0.0056	Actinobacteria	0.0176	Pseudomonas	0.0764
5	Firmicutes	0.0049	Planctomycetacia	0.0055	Novosphingobium	0.0501
6	Deinococcus-Thermus	0.0044	Deinococci	0.0044	Caulobacter	0.0464
7	Aquificae	0.0034	Deltaproteobacteria	0.0042	Phenylobacterium	0.0363
8	Verrucomicrobia	0.0018	Bacilli	0.0035	Sediminibacterium	0.0289
9	Acidobacteria	0.0011	Aquificae	0.0034	Sphingobium	0.0270
10	Chlamydiae	0.0009	Verrucomicrobiae	0.0018	Bosea	0.0227

Frequency Number of sequences of each phylogenetic level / total number of sequences (8,398,717).

Moreover, we determined the differences in the microbiota between samples in which *Legionella* spp. were present (positive samples) or absent (negative samples), based on both 16S and 18S rRNA sequencing and culture methods, to identify the microorganisms that coexisted with *Legionella* spp.

RESULTS AND DISCUSSION

Sequencing and taxonomic assignment

A total of 9,429,223 high-quality 16S rRNA sequences were obtained from the 112 bathwater samples after Illumina MiSeq sequencing, quality filtering, and chimera screening. The average, maximum, and minimum numbers of 16S rRNA sequences in each sample were 84,189; 190,987; and 20,733; respectively. Additionally, 12,179,023 high-quality 18S rRNA sequences were obtained. The average, maximum, and minimum numbers of 18S rRNA sequences in each sample were 108,741; 221,475; and 31,435; respectively.

After conducting BLASTn ver. 2.8.1[17] searches against the SILVA ver. 132 database[18], we examined the taxonomic information of sequences within the threshold values. Taxonomic information was assigned to 8,398,717 sequences. The average assignment rate for each sample was 89.1% (average 74,989; maximum 144,333; and minimum 16,389 sequences among the 112 samples). In addition, BLASTn analysis was performed on 18S rRNA sequences. A total of 1,867,244 18S rRNA sequences were assigned, and the average assignment rate for each sample was 15.3% (average 16,672, maximum 172,578, and minimum 11 sequences). Many samples presented a low percentage of assignments in 18S rRNA sequences, as these samples had a large number of assignments with prokaryotes.

Tables 1 and 2 show the top 10 frequencies of the number of sequences per phylogenetic group relative to the total number of sequences identifying the genus using 16S rRNA (8,398,717 sequences) and 18S rRNA (1,867,244 sequences) genes. Proteobacteria was the most frequently detected phylum in 16S rRNA sequences. The most common class was Alphaproteobacteria, followed by Gammaproteobacteria, both of which accounted for approximately 90% of all classes. At the genus level, we detected *Methylobacterium*, *Sphingomonas*, *Acinetobacter*, and *Pseudomonas*, which are widely found in the environment. Based on 18S rRNA sequences, eight groups were detected as supergroups. Among them, Opisthokonta, Amoebozoa, and SAR accounted for more than 90% of the total supergroups. *Vermamoeba*, a known *Legionella* spp. host[19,20], was the most frequently detected genus, accounting for approximately 25% of the total genera.

Legionella spp. detection using culture method and 16S rRNA sequencing

We compared the detection rates of *Legionella* spp. using the culture method and 16S rRNA sequencing. The culture method indicated that 72 of the 112 samples were positive (detection rate 68.8%, average detection rate 821 CFU per 100 mL, maximum detection rate 18,800 CFU per 100 mL, and minimum detection rate 0 CFU per 100 mL). The 16S rRNA sequencing detected *Legionella* spp. in 86 of the 112 samples (detection rate 76.8%, total number of detected sequences 10,270, and relative abundance of total sequences 0.0012) (10,270/8,398,717 sequences). The average and maximum number of sequences per sample was 92 and 1268, respectively, whereas the average and maximum relative abundance per sample was 0.0012 and 0.0144, respectively (Supplementary Table 1). The relative abundance of *Legionella* spp. per sample was higher than the mean relative abundance of all the identified genera (0.0008). Figure 1 shows a scatter plot of the *Legionella* spp. content obtained using the culture method (CFU per 100 mL) and the number of reads assigned to *Legionella* spp. using 16S rRNA sequencing in the 112 samples. The Spearman's rank correlation coefficients between the two methods were almost identical to those obtained in a previous study that compared culture and real-time PCR methods for *Legionella* spp. detection in bathwater. This result is in agreement with the results of Guillemet et al.[21] and Bontta et al.[22]. They reported a significant but weak correlation between the concentrations of *Legionella* spp. obtained using real-time PCR and those obtained using conventional culture methods in water samples.

Subsequently, we compared the detection and non-detection status of *Legionella* spp. using the culture method and 16S rRNA sequencing (Table 3, sample classification detailed in the Methods). The agreement between the detection status of the culture method and 16S rRNA sequencing was relatively high: $D/P = 87.8\%$, $ND/N = 53.3\%$. In particular, a high degree of conformity was observed between the samples that tested positive for *Legionella* spp. in the culture method and those that tested positive in 16S rRNA sequencing (D/P). These results showed that the quantitative correlation between both methods was low, but the qualitative correlation was high. On the contrary, the agreement between the samples that tested negative in the culture method and positive in 16S rRNA sequencing (D/N) was relatively high at 46.7%. This pattern may be due to the difference

Table 2. Top 10 frequencies of the phylogenetic groups identified by 18S rRNA sequences.

18S rRNA				
	Supergroup		Genus	
Ranking	Name	Frequency	Name	Frequency
1	Opisthokonta	0.383	Vermamoeba	0.2416
2	Amoebozoa	0.289	Trichosporon	0.0911
3	SAR	0.265	Debaryomyces	0.0577
4	Archaeplastida	0.037	Vishniacozyma	0.0492
5	Excavata	0.017	Epistylis	0.0370
6	Cryptista	0.008	Chromulinales JBNA46	0.0369
7	Haptista	0.0009	Filobasidium	0.0320
8	Incertae Sedis	0.0001	Rhodotorula	0.0265
9	-	-	Bromeliothrix	0.0251
10	-	-	Malassezia	0.0197

Frequency Number of sequences of each phylogenetic level / total number of sequences (1,867,244).

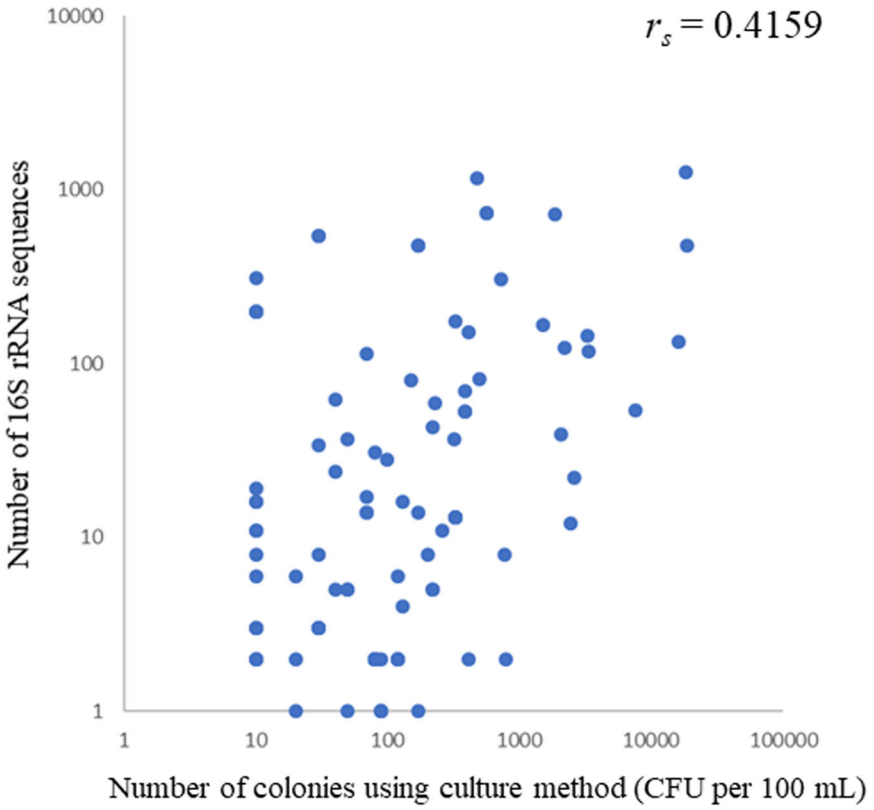

Fig. 1 Scatter plot of *Legionella* spp. content obtained using the culture method and the number of reads assigned using the 16S rRNA sequencing in the 112 samples. Each axis is shown in logarithmic scale. r_s is the Spearman's rank correlation coefficient between the numbers of 16S rRNA sequences and colonies in the 112 samples.

in the measurement principle of the two methods. In contrast to the real-time PCR method, the culture method detects only viable and culturable *Legionella* spp. and underestimates *Legionella* spp. present in protozoa[21]. The agreement between samples that tested positive in the culture method and negative in 16S rRNA sequencing (*ND/P*) was also relatively high at 12.2%. This pattern might be explained by disregarding PCR inhibitors in the extracted DNA[21]; alternatively, it could be due to the presence of PCR biases during amplicon sequencing. Further analysis of this cause is needed in the future.

In this study, we found a weak correlation in *Legionella* spp. abundance between the culture method and 16S rRNA sequencing. On the contrary, when we consider only the qualitative results (detection and non-detection), the degree of consistency between the culture method and 16S rRNA sequencing was high. The growth of *Legionella* spp. is considered to be subject to complex interactions with the habitat, such as water quality and coexisting microorganisms[23]. To elucidate the presence of microbes coexisting with *Legionella* spp. as a first step, we compared the samples using the presence or absence of each microbe as an indicator.

Table 3. *Legionella* spp. testing results using the culture method and 16S rRNA sequencing in 112 samples.

		16S rRNA Metagenomic analysis	
		Detected (*D*)	Not detected (*ND*)
Culture test	Positive (*P*)	72	10
		87.8%	12.2%
	Negative (*N*)	14	16
		46.7%	53.3%

The number of samples is indicated in the upper row, and the percentage of the 112 samples is shown in the lower row.

Table 4. Number of positive and negative samples.

Category	Number of samples
P_D +	22
P_D	50
P_ND	10
N_D	14
N_ND	16

P Samples positive for *Legionella* spp. in the culture method. *N* Samples negative for *Legionella* spp. in the culture method. *D* Samples positive for *Legionella* spp. using 16S rRNA sequencing. *ND* Samples negative for *Legionella* spp. using 16S rRNA sequencing. *P_D +* Samples whose *Legionella* spp. relative abundance was greater than the average relative abundance of all detected microbial genera (positive samples). *N_ND* Negative samples.

Microbiota comparison between the positive and negative groups

The 112 samples were classified into four groups: P_D, P_ND, N_D, and N_ND. The P_D + (positive group) included 22 samples, whose *Legionella* spp. relative abundance per sample was greater than the average relative abundance (0.001) of all detected microbial genera. In contrast, 16 samples were classified as N_ND (negative group) (Supplementary Table 1). The classification of each sample is shown in Table 4.

A total of 1,343 genera (956 prokaryotes and 387 eukaryotes) were identified in the 38 positive and negative samples (Supplementary Table 2). Table 5 shows the top genera with high detection frequency (the presence or absence of a species in a sample) in the P_D + and N_ND groups. We observed that some genera were common to both P_D + and N_ND groups, whereas others were unevenly distributed. This pattern indicates that the microbiota may differ between the two groups.

To confirm the differences in microbiota between the P_D + and N_ND groups, we performed multidimensional scaling (MDS) analysis using the Jaccard dissimilarity indexes for detection frequency in the positive and negative samples of all 1,343 genera (Fig. 2). The analysis of similarities (ANOSIM) showed a statistic R of 0.777 ($p = 0.001$). The positive and negative samples were classified into different clusters, thereby confirming the differences in microbiota between the P_D + and N_ND groups.

Extraction of microbial genera coexisting with *Legionella* spp.

Figure 3 shows the relationship between the number of occurrences of each of the 1343 genera detected in the positive and negative samples. The genera that appeared in only one sample accounted for approximately 40% of the total, suggesting that most microorganisms do not show any association with *Legionella* spp. and are likely rare microorganisms present in the samples. The prokaryotes *Acinetobacter*[24], *Flavobacterium*[25], *Methylobacterium*[26], *Pseudomonas*[27], and *Sphingomonas*[28] are commonly found in the environment, including drinking water, groundwater, and soil (Table 1). These prokaryotes might be universally present in the samples regardless of the presence of *Legionella* spp.

To extract only the microorganisms associated with *Legionella* spp., we removed the microorganisms that were rarely and commonly detected in the positive and negative samples. A total of 1224 (approximately 90% of 1343 genera) and 44 genera were classified as rare and common microorganisms, respectively (detailed in the Methods). After removing these microorganisms, 75 genera were extracted. The classification of each genus is shown in Supplementary Table 2.

We performed MDS analysis using the Jaccard dissimilarity indexes for the detection frequency of these 75 genera (Fig. 4). The ANOSIM statistic R was 0.962 ($p = 0.001$), clearly showing the differences in microbiota between the P_D + and N_ND groups, compared with that before microbial removal ($R = 0.777$). Next, we determined the Spearman's rank correlation coefficients against detection frequency in positive and negative samples between the 75 extracted genera and *Legionella* spp., and the top 10 genera are listed in Table 6. The Spearman's rank correlation coefficients between *Legionella* spp. and 1343 genera are shown in Supplementary Table 2.

Among prokaryotes, *Methyloversatilis* ($r_s = 0.89$), *Cupriavidus* ($r_s = 0.85$), and *Phenylobacterium* ($r_s = 0.84$) had high correlation coefficients. Among eukaryotes, *Vermamoeba* ($r_s = 0.77$) and *Aspidisca* ($r_s = 0.58$) were highly correlated with *Legionella* spp. *Methyloversatilis*, which had the highest correlation coefficient in prokaryotes, has been reported as a microorganism that forms biofilms in drinking water pipes and serves as a food source for amoebae[29]. Moreover, *Methyloversatilis*, *Phenylobacterium*, and *Caulobacter* have been detected together with *Legionella* spp. in biofilms formed in water pipes in artificial environments[30]. Similarly, *Reyranella* and *Bosea* have been isolated from biofilms in tap water by coculture with amoebae[23]. *Vermamoeba*, which had a high correlation in eukaryotes, is considered a *Legionella* spp. host and has been detected together with *Legionella* spp. in household hoses[31,32]. *Aspidisca* has also been detected in drinking water samples[33].

Contrary to expectations, *Acanthamoeba* and *Naegleria*, which are protists commonly associated with *Legionella* spp.[34], were not highly correlated in this study. Indeed, *Acantamoeba* was classified as a rare microorganism. *Naegleria* was extracted as one of the 75 genera highly related to *Legionella* spp. However, *Naegleria* was detected in a large number of samples that tested negative in the culture method, resulting in a low correlation with *Legionella* spp. The interaction between *Legionella* spp. and protozoa has been previously reported[34], and it can be influenced by a number of factors; the identity of the host cell, variations in the predatory behavior or feeding preferences of the host, the strain or species of the bacterium, the relative abundance of the two organisms, the external environment, and other microorganisms, which may have led to the present results[34]. This result may be partially due to the low hit rate (15.3%) against the 18S rRNA sequence database; however, to what extent this skews the genus identification remains unclear.

Several previous studies have determined the microbiota of artificially formed biofilms using non-chlorinated water[29,30]. In contrast, previous studies have reported that there is a predominant difference in the amount of *Legionella* spp. present in water at the boundary of $\geq 0.2\ mg\ L^{-1}$ residual chlorine in the water[35,36]. However, as *Legionella* spp. survive and multiply in biofilms, *Legionella* spp. suspended in water are disinfected by chlorine, whereas those present in the microflora in biofilms are able to survive as a result of the high resistance of the biofilm to disinfectants[20,34]. We used bathwater samples, which are highly likely to contain chlorine. The similarity of the microbiota detected, regardless of the habitat of *Legionella* spp., suggests

Table 5. Genera with high detection frequency in the positive and negative samples.

Domein	Class	Genus	Number of detected samples		Genus detected with high frequency	
			P_D +	N_ND	P_D +	N_ND
Bacteria	Alphaproteobacteria	Methylobacterium	22	16	+	+ +
Bacteria	Alphaproteobacteria	Sphingomonas	22	15	+	+ +
Bacteria	Gammaproteobacteria	Legionella	22	0	+	
Bacteria	Gammaproteobacteria	Acinetobacter	22	15	+	+ +
Bacteria	Gammaproteobacteria	Pseudomonas	22	15	+	+ +
Bacteria	Bacteroidia	Flavobacterium	21	16	+	+ +
Bacteria	Alphaproteobacteria	Caulobacter	21	5	+	
Bacteria	Alphaproteobacteria	Novosphingobium	21	9	+	
Bacteria	Alphaproteobacteria	Sphingopyxis	21	8	+	
Bacteria	Gammaproteobacteria	Methyloversatilis	21	1	+	
Bacteria	Gammaproteobacteria	Limnohabitans	18	16		+ +
Eukaryota	Basidiomycota	Malasseziales	13	15		+ +
Eukaryota	Oligohymenophorea	Ophrydium	11	15		+ +
Eukaryota	Ochrophyta	Paraphysomonas	8	16		+ +
Bacteria	Gammaproteobacteria	Polynucleobacter	3	16		+ +
Bacteria	Actinobacteria	Rhodoluna	2	16		+ +
Bacteria	Bacteroidia	Pseudarcicella	2	16		+ +

Genera appearing in 22 or 21 detections in the negative samples are marked with +, whereas genera appearing in 16 or 15 detections in the negative samples are marked with + +.

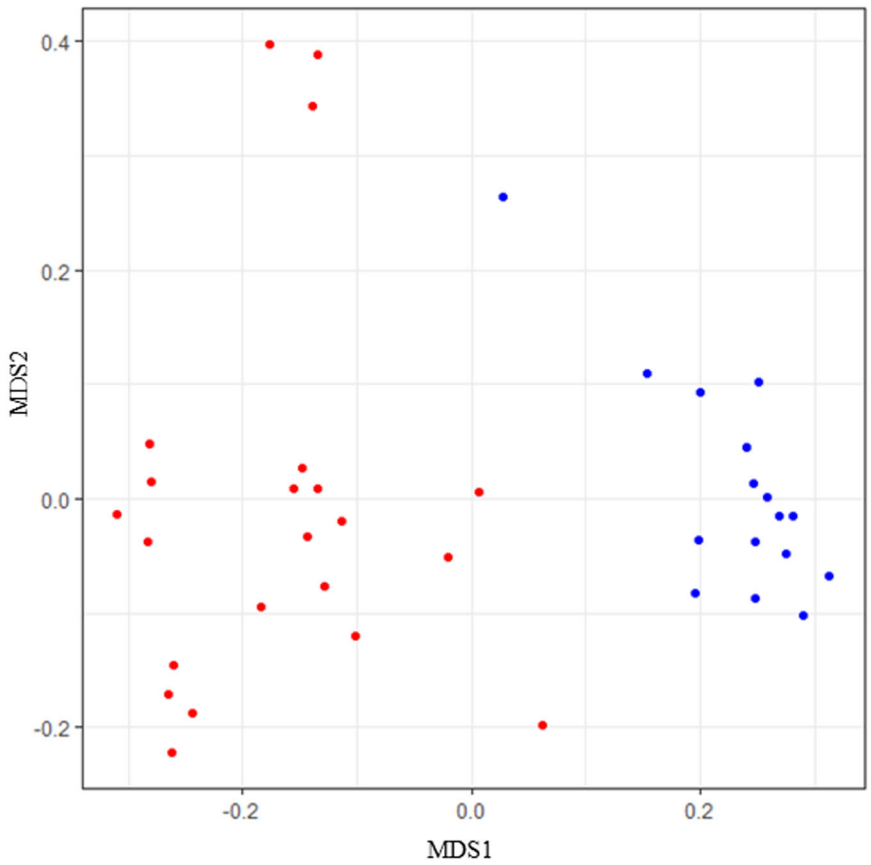

Fig. 2 Multidimensional scaling (MDS) plot of 956 prokaryotic and 387 eukaryotic genera. Red dots show positive samples ($N = 22$) and blue dots show negative samples ($N = 16$).

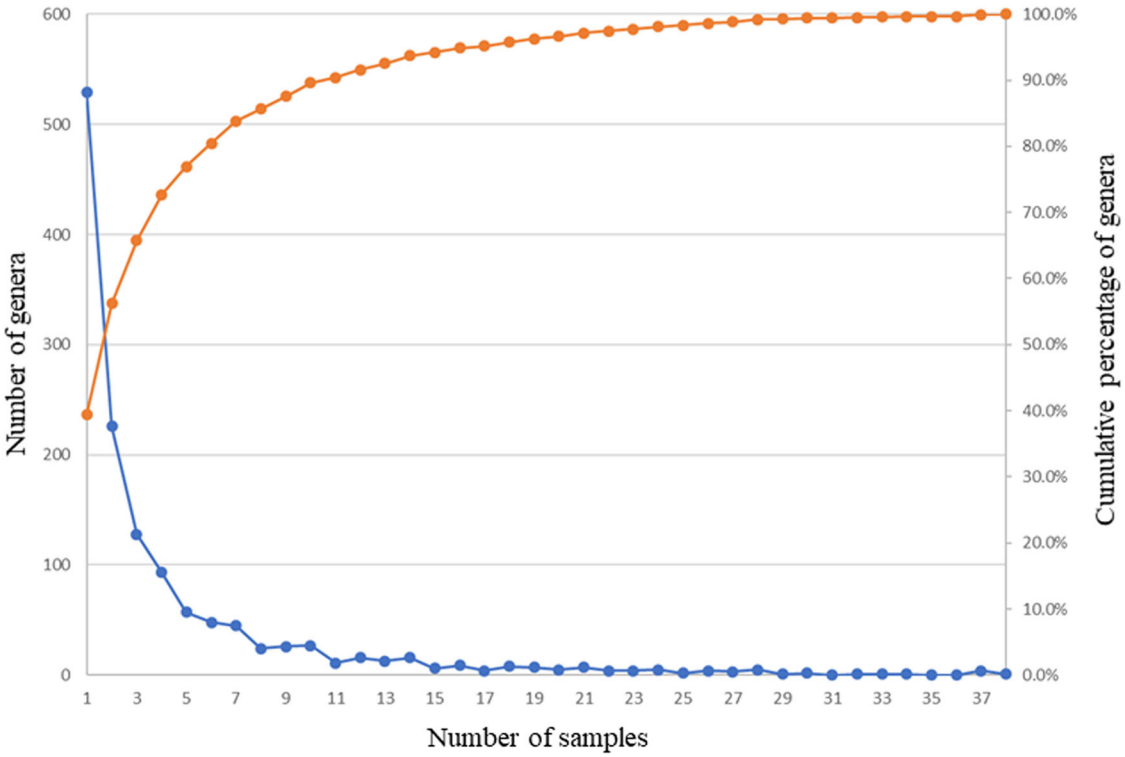

Fig. 3 Number (blue line) and cumulative percentage (orange line) of genera commonly detected within the positive and negative samples. The cumulative percentage was calculated as cumulative genera/1343 genera × 100.

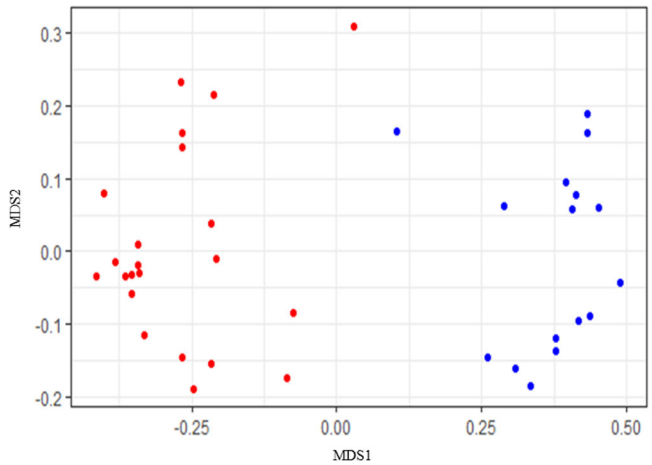

Fig. 4 Multidimensional scaling (MDS) plot of 75 selected genera. Red dots show positive samples ($N = 22$), and blue dots show negative samples ($N = 16$).

Table 6. Genera associated with the presence of *Legionella* spp.

16S rRNA		18S rRNA	
Genus	r_s	Genus	r_s
Methyloversatilis	0.89	Vermamoeba	0.77
Cupriavidus	0.85	Aspidisca	0.58
Phenylobacterium	0.84	Arcuospathidium	−0.28
Kinneretia	0.73	Ochromonas	−0.34
Ferrovibrio	0.70	Epistylis	−0.38
Caulobacter	0.68	Bromeliophrya	−0.40
Reyranella	0.67	Cercomonas	−0.40
Pseudorhodoplanes	0.65	Naegleria	−0.40
Mesorhizobium	0.65	Paramecium	−0.45
Bosea	0.62	Tetrahymena	−0.47

r_s Spearman's rank correlation coefficient against detection frequency in positive and negative samples between the genus and *Legionella* spp.

that the microorganisms extracted in this study may be closely related to the survival and growth of *Legionella* spp.

In conclusion, we observed differences in the microbiota of the 112 bathwater samples with and without *Legionella* spp. The prokaryotes *Methyloversatilis*, *Cupriavidus*, and *Phenylobacterium* and the eukaryotes *Vermamoeba* and *Aspidisca* were highly correlated with *Legionella* spp. In terms of the habitat of *Legionella* spp., most previous studies were based on the microbiota in environments where residual chlorine was not present, whereas the results of this study were obtained in environments where residual chlorine was likely to be present. Therefore, it is suggested that *Legionella* spp. develop in a specific microbiota, regardless of the habitat or formation process. However, as the chlorine concentration was not actually measured in this study, it is necessary to accurately determine the presence or absence of residual chlorine in the future. In addition, other habitats, such as cooling towers, should be analyzed to clarify the composition of the microbiota coexisting with *Legionella* spp.

METHODS

Sample collection

A total of 112 independent bathwater samples were collected from bathtubs in bathing facilities in Japan between February 2016 and November 2018. Details of the collection date for each sample are provided in Supplementary Table 1. All samples were collected in

200 mL polyethylene flasks containing sterile sodium thiosulfate and stored in the dark at approximately 4 °C until testing.

Legionella spp. detection via the culture method

In this study, *Legionella* spp. were detected using the filtration method commonly used in Japan[9]. Each water sample (200 mL) was concentrated using filtration through a 0.2 µm pore size membrane filter (Advantec Tokyo Co., Ltd., Tokyo, Japan). Next, the membrane was immersed in 4 mL of sterile distilled water and vortexed for 2 min. The suspension (1 mL) was supplemented with 1 mL of 0.2 M HCl-KCl buffer (pH 2.2), heated for 5 min at approximately 25 °C, and incubated on two Wadowsky-Yee-Okuda agar plates (100 µL per plate) supplemented with α-ketoglutarate (Eiken Chemical Co., Ltd., Tokyo, Japan) for 5–7 days at 36 ± 1 °C. After incubation, colonies (1−10) with *Legionella* spp. were cultured on *Legionella* buffered charcoal yeast extract (BCYE-α) agar (Nikken Bio Co., Ltd., Tokyo, Japan). After 3 days of incubation at 36 ± 1 °C, the number of colonies per 100 mL was determined for isolates that grew on BCYE-α but not on blood agar base agar. The remaining suspension was stored at −20 °C and used for DNA extraction. The detection limit of the culture method is 1 colony per 10 mL, i.e., 10 colonies per 100 mL of bathwater.

Sequencing of 16S and 18S rRNA genes and bioinformatics methods

Mixed genomic DNA from each sample was extracted using the NucleoSpin Microbial DNA kit (Macherey-Nagel, Germany) and was used to amplify the bacterial 16S rRNA and eukaryotic 18S rRNA regions using PCR. Primers 515 F and 806 R amplified the 16S rRNA V4 region, whereas primers 1389 F and 1510 R amplified the 18S rRNA V9 region. Sequencing of 16S and 18S rRNA genes was performed by Fasmac Co., Ltd. (Atsugi, Japan) using the Illumina MiSeq sequencing platform. The 16S and 18S rRNA sequences were deposited in the DNA Data Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp) Sequence Read Archive under accession no. DRA014009.

To assign taxonomic information using sequence similarity, we performed BLASTn ver. 2.8.1[17] searches against the SILVA ver. 132 database[18] using 16S and 18S rRNA sequences. We examined the taxonomic information of sequences showing the best hit with an *E*-value of ≤ 1e-5 and both sequence identity and coverage were ≥ 97%. If the 16S and 18S rRNA sequences were similar to a reference sequence with an unidentified genus, the sequences were removed.

Sample classification

To identify samples in which *Legionella* spp. are reliably detected, samples were labeled based on the detection of *Legionella* spp. using the culture method and 16S rRNA sequencing. First, samples in which *Legionella* spp. were detected using the culture method were defined as "*P*," whereas samples in which *Legionella* spp. were not detected were defined as "*N*." Next, samples in which *Legionella* spp. were detected using 16S rRNA sequencing were defined as "*D*," whereas samples in which *Legionella* spp. was not detected were defined as "*ND*." By combining the labels of the culture method and 16S rRNA sequencing, the samples were classified into four groups: P_D, P_ND, N_D, and N_ND. Among P_D samples, those with *Legionella* spp. relative abundance per sample greater than the average relative abundance of all detected microbial genera were defined as P_D +. In this study, P_D + was defined as the positive and N_ND as the negative group.

Microbiota comparison in the positive and negative groups

To compare the microbial profiles of the P_D + and N_ND groups, we performed MDS using the Jaccard dissimilarity index, which was calculated from the genus presence-absence matrix in the R package vegan[37]. ANOSIM is a non-parametric test that was used to compare the differences between and within groups based on the ranks of the Jaccard dissimilarity distances[38]. The ANOSIM statistic (R) was calculated by comparing the mean of all rank Jaccard dissimilarities for samples between and within groups. The R values range from −1 to 1; values close to 1 indicate high dissimilarity between groups, whereas values close to −1 indicate high dissimilarity within groups. A value of zero indicates completely random dissimilarity[38,39].

Elimination of rarely and commonly detected microorganisms in the positive and negative groups

In 16S and 18S rRNA sequencing using next-generation sequencers, the majority of the microorganisms assigned by one or two sequences may be due to sequence errors, and microorganisms with low numbers are typically excluded from the analysis[40–42]. In addition, common microorganisms detected in both P_D + and N_ND groups were considered as indigenous bacteria in the bathwater. The elimination criteria were defined to remove rarely and commonly detected microorganisms in both P_D + and N_ND groups. The detection frequency of a microorganism (X) in the P_D + group (SP_x) was calculated as $SP_x = X_p/P$; where, X_p is the number of microorganisms (X) detected in the P_D + group and P is the number of positive samples. In addition, the detection frequency of a microorganism (X) in the N_ND group (SN_x) was calculated as $SN_x = X_n/N$; where, X_n is the number of microorganisms (X) detected in the N_ND group and N is the number of negative samples.

Microorganisms with SP_x and $SN_x < 0.5$ were defined as rare microorganisms and removed from both P_D + and N_ND groups. Next, microorganisms with detection frequency ratios (SP_x/SN_x) ≥ 0.5 and ≤ 2 were defined as commonly detected microorganisms and removed from both P_D + and N_ND groups.

Statistical analysis

All statistical analyses, including Spearman's rank correlation analysis, were performed using the R package[43].

DATA AVAILABILITY

The 16S and 18S rRNA sequences were deposited in the DNA Data Bank of Japan (DDBJ) Sequence Read Archive under accession no. DRA014009. The data generated and/or analyzed in this study are available from the corresponding author on reasonable request.

CODE AVAILABILITY

All codes used in this study are available from the corresponding author on reasonable request.

Received: 17 June 2022; Accepted: 15 December 2022;
Published online: 24 December 2022

REFERENCES

1. Mercante, J. W. & Winchell, J. M. Current and emerging *Legionella* diagnostics for laboratory and outbreak investigations. *Clin. Microbiol. Rev.* **28**, 95–133 (2015).
2. Tanimoto, T., Takahashi, K. & Crump, A. Legionellosis in Japan: A self-inflicted wound? *Intern. Med.* **60**, 173–180 (2021).
3. Declerck, P. Biofilms: The environmental playground of *Legionella pneumophila*. *Environ. Microbiol.* **12**, 557–566 (2010).
4. Parte, A. C. LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* **68**, 1825–1829 (2018).
5. Yu, V. L. et al. 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 (2002).
6. Phin, N. et al. Epidemiology and clinical management of Legionnaires' disease. *Lancet Infect. Dis.* **14**, 1011–1021 (2014).

7. Kuroki, T., Ishihara, T., Ito, K. & Kura, F. Bathwater-associated cases of legionellosis in Japan, with a special focus on *Legionella* concentrations in water. *Jpn. J. Infect. Dis.* **62**, 201–205 (2009).
8. Villari, P., Motti, E., Farullo, C. & Torre, I. Comparison of conventional culture and PCR methods for the detection of *Legionella pneumophila* in water. *Lett. Appl. Microbiol.* **27**, 106–110 (1998).
9. Edagawa, A. et al. Detection of culturable and nonculturable *Legionella* species from hot water systems of public buildings in Japan. *J. Appl. Microbiol* **105**, 2104–2114 (2008).
10. Joly, P. et al. Quantitative real-time *Legionella* PCR for environmental water samples: Data interpretation. *Appl. Environ. Microbiol.* **72**, 2801–2808 (2006).
11. Ta, A. C., Stout, J. E., Yu, V. L. & Wagener, M. M. Comparison of culture methods for monitoring *Legionella* species in hospital potable water systems and recommendations for standardization of such methods. *J. Clin. Microbiol.* **33**, 2118–2123 (1995).
12. Boulanger, C. A. & Edelstein, P. H. Precision and accuracy of recovery of *Legionella pneumophila* from seeded tap water by filtration and centrifugation. *Appl. Environ. Microbiol.* **61**, 1805–1809 (1995).
13. Touron-Bodilis, A., Pougnard, C., Frenkiel-Lebossé, H. & Hallier-Soulier, S. Usefulness of real-time PCR as a complementary tool to the monitoring of *Legionella* spp. and *Legionella pneumophila* by culture in industrial cooling systems. *J. Appl. Microbiol* **111**, 499–510 (2011).
14. Steinert, M., Emödy, L., Amann, R. & Hacker, J. Resuscitation of viable but nonculturable *Legionella pneumophila* Philadelphia JR32 by Acanthamoeba castellanii. *Appl. Environ. Microbiol.* **63**, 2047–2053 (1997).
15. Peabody, M. A. et al. Characterization of *Legionella* Species from Watersheds in British Columbia, Canada. *mSphere* **2**, e00246–17 (2017).
16. Llewellyn, A. C. et al. Distribution of *Legionella* and bacterial community composition among regionally diverse US cooling towers. *PLoS One* **12**, e0189937 (2017).
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
18. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
19. Harb, O. S., Gao, L. Y. & Abu Kwaik, Y. From protozoa to mammalian cells: a new paradigm in the life cycle of intracellular bacterial pathogens. *Environ. Microbiol.* **2**, 251–265 (2000).
20. Lau, H. Y. & Ashbolt, N. J. The role of biofilms and protozoa in *Legionella* pathogenesis: implications for drinking water. *J. Appl. Microbioll.***107**, 368–378 (2009).
21. Guillemet, T. A. et al. Assessment of real-time PCR for quantification of *Legionella* spp. in spa water. *Lett. Appl. Microbiol.* **51**, 639–644 (2010).
22. Bonetta, S., Bonetta, S., Ferretti, E., Balocco, F. & Carraro, E. Evaluation of *Legionella pneumophila* contamination in Italian hotel water systems by quantitative real-time PCR and culture methods. *J. Appl. Microbiol.* **108**, 1576–1583 (2010).
23. Paranjape, K. et al. Unravelling the importance of the eukaryotic and bacterial communities and their relationship with *Legionella* spp. ecology in cooling towers: a complex network. *Microbiome* **8**, 157 (2020).
24. Baumann, P. Isolation of *Acinetobacter* from soil and water. *J. Bacteriol.* **96**, 39–42 (1968).
25. Michael, E. B. & Michael, L. Brown. A review of *Flavobacterium psychrophilum* biology, clinical signs, and bacterial cold water disease prevention and treatment. *Open Fish. Sci. J.* **4**, 40–48 (2011).
26. Gallego, V., García, M. T. & Ventosa, A. *Methylobacterium isbiliense* sp. nov., isolated from the drinking water system of Sevilla, Spain. *Int. J. Syst. Evol. Microbiol.* **55**, 2333–2337 (2005).
27. Vaz-Moreira, I., Nunes, O. C. & Manaia, C. M. Diversity and antibiotic resistance in *Pseudomonas* spp. from drinking water. *Sci. Total Environ.* **426**, 366–374 (2012).
28. White, D. C., Sutton, S. D. & Ringelberg, D. B. The genus *Sphingomonas*: Physiology and ecology. *Curr. Opin. Biotechnol.* **7**, 301–306 (1996).
29. van der Kooij, D., Veenendaal, H. R., Italiaander, R., van der Mark, E. J. & Dignum, M. Primary colonizing Betaproteobacteriales play a key role in the growth of *Legionella pneumophila* in biofilms on surfaces exposed to drinking water treated by slow sand filtration. *Appl. Environ. Microbiol.* **84**, e01732–18 (2018).
30. Proctor, C. R., Reimann, M., Vriens, B. & Hammes, F. Biofilms in shower hoses. *Water Res.* **131**, 274–286 (2018).
31. Valster, R. M., Wullings, B. A. & van der Kooij, D. Detection of protozoan hosts for *Legionella pneumophila* in engineered water systems by using a biofilm batch test. *Appl. Environ. Microbiol.* **76**, 7144–7153 (2010).
32. Thomas, J. M., Thomas, T., Stuetz, R. M. & Ashbolt, N. J. Your garden hose: a potential health risk due to *Legionella* spp. growth facilitated by free-living amoebae. *Environ. Sci. Technol.* **48**, 10456–10464 (2014).
33. Valster, R. M., Wullings, B. A., Bakker, G., Smidt, H. & van der Kooij, D. Free-living protozoa in two unchlorinated drinking water supplies, identified by phylogenic analysis of 18S rRNA gene sequences. *Appl. Environ. Microbiol.* **75**, 4736–4746 (2009).
34. Boamah, D. K., Zhou, G., Ensminger, A. W. & O'Connor, T. J. From many hosts, one accidental pathogen: The diverse protozoan hosts of *Legionella*. *Front Cell Infect. Microbiol.* **7**, 477 (2017).
35. Qin, T. et al. High prevalence, genetic diversity and intracellular growth ability of *Legionella* in hot spring environments. *PLoS One* **8**, e59018 (2013).
36. Mouchtouri, V. et al. Risk factors for contamination of hotel water distribution systems by *Legionella* species. *Appl. Environ. Microbiol.* **73**, 1489–1492 (2007).
37. Oksanen, J. et al. Package 'vegan'. Retrieved from http://CRAN.R-project.org/package=vegan (2020).
38. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
39. Clarke, K. R. & Green, R. H. Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Prog. Ser.* **46**, 213–216 (1988).
40. Baltar, F. et al. Response of rare, common and abundant bacterioplankton to anthropogenic perturbations in a Mediterranean coastal site. *FEMS Microbiol. Ecol.* **91**, fiv058 (2015).
41. Kerrigan, Z. & D'Hondt, S. Patterns of relative bacterial richness and community composition in seawater and marine sediment are robust for both operational taxonomic units and amplicon sequence variants. *Front. Microbiol.* **13**, 796758 (2022).
42. Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–123 (2010).
43. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/ (2021).

ACKNOWLEDGEMENTS

A part of the work described in this report was funded by a grant from Shiga Prefecture Subsidy for New Technology Development Project for Small and Medium Enterprises. We would like to thank Editage (www.editage.com) for English language editing.

AUTHOR CONTRIBUTIONS

T.A. and A.K. designed the research. M.O., K.K., A.K., and T.A. performed the research. R.H. and A.K. contributed new materials. M.O., K.K., and T.A. analyzed the data. M.O. drafted the manuscript. All authors revised the paper, participated in discussions throughout the research, and approved the final version of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41545-022-00212-y.

Correspondence and requests for materials should be addressed to Takashi Abe.

Reprints and permission information is available at http://www.nature.com/reprints

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