Structure of bacterial communities in Japanese-style bathrooms: Comparative sequencing of bacteria in shower water and showerhead biofilms using a portable nanopore sequencer

So Fujiyoshi[a,b,c,d*], Yukiko Nishiuchi[a], Fumito Maruyama[a,b,c,d**]

a. *Office of Industry-Academia-Government and Community Collaboration, Hiroshima University, Hiroshima, Japan.*

b. *Center for Holobiome and Built Environment (CHOBE), Hiroshima University, Hiroshima, Japan.*

c. *Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan.*

d. *Network for Extreme Environment Research (NEXER), Scientific and Technological Bioresource Nucleus (BIOREN), Universidad de La Frontera, Temuco, Chile.*

***Corresponding author**: So Fujiyoshi, Office of Industry-Academia-Government and Community Collaboration, Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8511, Japan. E-mail: fujiyoshi.so.62w@kyoto-u.jp, Tel: +81(82)424-7048, Orchid ID: 0000-0003-4306-332X.

****Co-corresponding author**: Fumito Maruyama, Office of Industry-Academia-Government and Community Collaboration, Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8511, Japan. E-mail: fumito@hiroshima-u.ac.jp, Tel: +81(82)424-7048, Orchid ID: 0000-0003-2347-616X.

Abstract

Showers are one of the main exposure routes to diverse microbes for end users in built environments. Bacteria in water are responsible for biofilm formation on surfaces, and the inside of a showerhead is a specific niche. Here, for the purpose of microbial characterization, source estimation and possibility of infection, the bacterial compositions of both shower water and showerhead biofilms in the same bathroom were determined and compared using a portable nanopore sequencer. The results suggest that specific bacteria in source water would primarily adhere to the surface of the showerhead where they subsequently form biofilms, and the community compositions within biofilms largely vary depending on environmental factors. The relative abundance of several pathogenic bacterial genera in both water and biofilm samples was low. We suggest that it is important to manage risk of infection in each household, and rapid on-site analysis of microbial communities will allow the realization.

Keywords

Built environment, Bathroom, Showerhead biofilm, Shower water, Nanopore sequencer, Comparative study, Bacterial community, Risk of infection

Introduction

We are surrounded by numerous microbes in our daily lives, and the bathroom is one of the largest reservoirs of microbes in the built environment due to moderately high temperatures and humidity (Neu *et al*. 2018). The typical Japanese bathing style is to immerse the body in the bathtub and wash the body outside the bathtub, and many Japanese people are known to bath for longer times and with higher frequency (approximately 30 min, every day) than people in other countries (Tochihara, 1999; Odamaki *et al*. 2019).

Shower water is generally supplied by a drinking-water distribution system, so microbes found in the water usually pose no risk for healthy individuals. To control microbial contamination in water, different chemicals and chemical treatment methods (ozone, fluorine, chlorine, chlorine dioxide, monochloramine, and copper-silver ionization and photocatalysis) and physical treatments (thermal inactivation, UV, filtration, and precipitation) are continuously applied (Cervero-Aragó *et al*. 2015; Huang *et al*. 2020). However, some of the microbes in the water can be opportunistic pathogens (Ops) capable of causing serious and life-threatening infections in severely immunocompromised individuals. Some Ops in water cause pneumonia (e.g., non-tuberculous mycobacteria (NTM), *Legionella* spp. and *Mycoplasma* spp.), asthma, or allergies (Dannemiller *et al*. 2016; Montagna *et al*. 2016; Nishiuchi *et al*. 2017). Bacteria in water are important in biofilm formation on surfaces, as they provide the initial cells for attachment and further biofilm development, and the inside of a showerhead is a specific niche that is moist, warm, dark, and frequently replenished with low-level nutrient resources. Bacteria are more concentrated in the biofilm than in the feed water; the number of colony forming units (CFUs) in water after chlorination is 10^2 CFU/mL, while in biofilms, the number of CFUs is between 10^4–10^5 CFU/mL (Kormas *et al*. 2010; Peng *et al*. 2020; Novak *et al*. 2020). In addition, biofilms provide protection against environmental stressors such as antimicrobial agents and disinfectants (Johnson 2008).

Previous microbiological studies of built environment have used culture methodology to detect and identify microbes and have focused primarily on *Legionella pneumophilia* and *Mycobacterium avium* subsp. *hominissuis* (MAH) (Arikawa *et al*. 2019; Falkinham 2020a). Nishiuchi *et al*. reported predominant colonization of *M. avium* in bathtub inlets of bathrooms of patients in Japan using culture isolation (Nishiuchi *et al*. 2007; Nishiuchi *et al*. 2009). Subsequently, Iwamoto *et al*. reported a high degree of genetic relatedness between bacterial isolates from pulmonary MAH patients and bacterial isolates from their bathrooms based on variable numbers of tandem repeats (VNTR) analysis of 19 loci (Iwamoto *et al*. 2012). These reports implied that bathrooms are potentially a major source of MAH in Japan. However, the main limitation of those studies was that it was unclear whether the transmission of the pathogen originated from the patient or the drinking water distribution systems. Thus, understanding the composition of the microbiome in both water and biofilms is of great importance for health risk assessment.

In 2012, a portable nanopore sequencer MinION (Oxford Nanopore Technologies, London, UK) was developed, which is now considered a breakthrough in DNA sequencing technology. The MinION sequencer enables real-time, on-site analyses of any genetic material (Schmidt *et al*. 2016; Parker *et al*. 2017; Nakagawa *et al*. 2019). Compared to the currently widely accepted MiSeq sequencer (Illumina, San Diego, CA, USA), the MinION sequencer can generate a much longer read length, although with lower accuracy. The increased information content inherent from longer read lengths helps researchers with alignment-based taxonomy assignments. Nygaard *et al*. analysed the building dust microbiome using MinION and MiSeq, and showed that MinION had a better taxonomic resolution than MiSeq at the genus and species levels. By optimizing the PCR conditions, MinION was shown to provide accurate microbial community composition and more accurate data than MiSeq in terms of species-level matches (Fujiyoshi *et al*. 2020). Thus, nanopore sequencing is an available and useful tool for understanding the status of public health based on microbial genetic information.

In this study, we used a high-throughput portable nanopore sequencer to determine the bacterial community in shower water and in showerhead biofilms.

Results

Sampling and Sequence reads

Fifty samples from 25 residential bathrooms (showerhead biofilm and shower water paired samples, hereafter called biofilm and water samples) were collected from the Hokkaido to Kinki regions in Japan (Figure 1A and 1B). As illustrated in Figure 2, microbes from biofilms were generally clumped and embedded in extracellular material, consistent with biofilm morphology. The DNA yields from the biofilm and water samples were highly variable, and detectable amounts of DNA could not always be extracted. After PCR, 7 biofilm or water samples were not amplified; therefore, 7 pairs of samples were removed for analysis of the pairs. After sequencing, 2 pairs of samples were removed due to a low number of sequences reads. Finally, 32 samples in 16 pairs were analysed in this study. Among the 643,780 raw reads from water samples and 489,734 raw reads from biofilm samples, 343,092 and 228,863 reads, respectively, were analysed after filtration (Supplemental Table S1). The mean operational taxonomic unit (OTU) counts from the water and biofilm samples were 400 and 279, respectively. The number of OTUs was not significantly different between water and biofilm samples (p-value =0.16 > 0.05). Good's coverage values were greater than 98% for all the samples (Supplemental Table S1).

Comparison of the bacterial community structure in water and biofilm samples

The taxon richness and evenness level reflected by the Shannon index was significantly higher in the water samples than in the biofilm samples ($p < 0.01$) (Figure 3 and Supplemental Table S1). In addition, the Chao1 index, which accounts for species richness, showed that phylogenetic diversity was significantly higher in water samples than in biofilm samples ($p < 0.01$) (Supplemental Table S1). The nMDS plots demonstrated that the bacterial communities in the water samples were more closely clustered than those in the biofilm samples (Figure 4). PERMANOVA results confirmed that the overall community structures were significantly different in both the water and biofilm samples ($p < 0.01$; see Figure 4 for F and R-squared values). Venn diagrams illustrate the percentage of genera shared among each pair of samples. In the biofilm samples, an average of 66.1% of the genera were common to water samples, whereas in the water samples, an average of 36.1% of genera were common to biofilm samples (Figure 5). The percentage of common genera in biofilms was significantly higher than that in water ($p < 0.01$), and the percentage of shared sequences was relatively higher in biofilms (84.1%) than in water (72.1%) ($p > 0.05$).

Bacterial community composition

The top three genera in the water samples were *Sphingomonas* (18.8%), *Methylobacterium* (17.0%), and *Phreatobacter* (12.6%), which belong to Proteobacteria (alpha-subclass) (Figure 6), while three genera, *Sphingomonas*, *Methylobacterium*, and *Bradyrhizobium,* were found in all the water samples. The top three genera in the biofilm samples were *Methylobacterium* (25.2%) and *Sphingomonas* (22.4%), belonging to the phylum Proteobacteria (alpha-subclass), and *Brevibacterium* (8.2%), belonging to the phylum Actinobacteria (Figure 6, for more detail information see Supplemental Figure S1). Only *Methylobacterium* was commonly found in all biofilm samples. To investigate the differences between shower water and biofilm samples, Linear discriminant analysis (LDA) effect size (LEfSe) was performed, and the bacterial taxa significantly associated with each sample was identified. A non-parametric factorial Kruskal-Wallis sum-rank test embedded in LEfSe identified 15 bacterial genera that were characteristic only for water but not for biofilm samples (Figure 7).

Risk of infection in bathroom environments based on the presence of pathogenic bacteria

Based on the BSL2 and BSL3 list from National Institute of Infectious Diseases (NIID), a total of 13 genera were identified in the water and biofilm samples (Figure 8). The genus present at the highest frequency was *Moraxella* with a prevalence of 81% (26/32) and a median value of 0.40%, and the genus with the lowest frequency was *Nocardia* with a prevalence of 3% (1/32) and a median of 0.00%. Only *Moraxella* and *Staphylococcus* had mean values above 1%, and both were detected at high rates in some residences. The median value was 0.00% for most genera. *Mycobacterium*, which includes Ops that cause pulmonary disease, was detected in 34% (11/32) of the samples, with a median of 0.00%. *Legionella* was not detected in this study. Even samples taken from the same bathroom had an Op genus present in the biofilm but not in the water, or vice versa. There was no significant difference in the frequency, mean or median of genera that include pathogens in the biofilm and water samples (p-value > 0.05). At the species level, 5 and 7 BSL2 or BSL3 species were found in biofilm and water samples, respectively. These species accounted for less than 0.30% of the total community (Supplemental Table S2). In addition, these species had a lower percent identity (80.4% average value of identity) when analysed against a more refined and updated BLASTn database (Supplemental Table S3). The species that appeared in both analyses were *Staphylococcus aureus* (87.98% maximum value of identity) and *Vibrio parahaemolyticus* (87.86% maximum value of identity). The genera that appeared in both analyses were *Corynebacterium* (86.26% maximum value of identity) and *Mycobacterium* (86.87% maximum value of identity). For the four species of *Brucella* found in this study, most of them were assigned to different genera with low percent identity (76.34% average value of identity).

Discussion

Biofilm bacteria are derived from the water

The biofilm formed on the inner surface of the shower head interacts with the shower water (Rasmus *et al*. 2002; Peng *et al*. 2020). The results of our Venn diagram also show that some genera are shared between biofilm and water samples (Figure 5). However, the bacterial community structures of biofilm and water samples were significantly different, even when the samples obtained from the same bathroom (Figures 3 and 4). The bacterial community in biofilms had a lower alpha-diversity than the community in the water samples, suggesting that some, but not all of the bacteria in the water probably adhere to the showerhead surface and subsequently form a biofilm. Beta diversity analysis revealed significant differences between water and biofilm samples (Figure 4). The bacterial communities in the water were more closely clustered than those in the biofilm samples. The quality of drinking water in Japan is standardized by the Ministry of Health, Labour and Welfare, based on a total of 51 items, such as general bacterial CFU, chlorine and cadmium concentration, and each standard has a maximum value below which health is not affected (Nishiuchi *et al*. 2017; Novak *et al*. 2020). All the water samples collected in this study were supplied from drinking water distribution systems (not well or river water), and the well-controlled water quality in Japan may affect the similarity of the bacterial communities in the water samples. However, the bacterial communities in the biofilms differed greatly among samples. This probably suggests that the bacterial community in a biofilm is influenced by factors other than the water, such as the building type, the method of use, and the residents (Ji *et al*. 2017; Proctor *et al*. 2018). In this study, there was no correlation between the material of the showerhead and the bacterial community structure (data not shown).

188 The bacterial communities examined in this study are consistent with other studies showing the
189 dominance of similar genera in freshwater systems and biofilms in bathrooms (Figure 6) (Feazel
190 *et al*. 2009). *Methylobacterium* spp. are famous for forming pink-coloured biofilms and are
191 prevalent in domestic water-associated environments, such as drinking water distribution
192 systems, shower curtains, and showerheads worldwide (Gallego *et al*. 2005; Zhou *et al*. 2021). In
193 Japan, *Methylobacterium* are also common bacteria with known culture methods (Yano *et al*.
194 2013; Kawai *et al*. 2019). *Sphingomonas*, generally the most abundant genus detected in the
195 biofilm samples, are ubiquitous in the environment, such as in soil, water, and sediments, as well
196 as on shower curtains and are known to be constantly and consistently present in biofilms even
197 after chlorination (White *et al*. 1996; Kelley *et al*. 2004; Douterelo *et al*., 2018). *Phreatobacter*
198 has been identified as a dominant genus in drinking water distribution systems in China (Li *et al*
199 2020; Ma *et al* 2020; Jing *et al*., 2021). *Brevibacterium* from biofilms are not as commonly
200 detected in bathrooms as the other three genera. *Brevibacterium* species are isolated from various
201 habitats, such as fermented food, animal and human skin, insects, soil, and mural paintings.
202 Some species, such as *B. casei* and *B. epidermidis*, encountered in human skin, wounds, and
203 blood can cause rare human infections, making them Ops (Denis and Irlinger, 2008). It should be
204 investigated in the future whether the high percentage of *Brevibacterium* in showerhead biofilms
205 is unique to Japan or due to environmental selective pressure. *Sphingomonas*, *Methylobacterium*,
206 and *Bradyrhizobium* were found in all the water samples, while only *Methylobacterium* was
207 found in all the biofilm samples. This shows that the bacteria in the water samples were
208 relatively consistent, but the biofilm constituent bacteria varied among samples. LEfSe analysis
209 identified significant differences in taxa between biofilm and water samples. A comparison of
210 sample sources indicated that there were 15 taxa only from the water samples with LDA scores
211 greater than 4.0 (Figure 7). No taxa identified were considered indicators of biofilms. This
212 suggests that the bacteria in the biofilm samples are of water origin. A total of 66.1% and 36.1%
213 of genera were common between the biofilm and water samples, respectively (Figure 5). These
214 results suggest that specific bacteria in the original water would primarily adhere to the surface
215 of the showerhead and subsequently form biofilms.

216 **Potentially pathogenic bacteria were present in shower water and biofilm samples, but only**
217 **a low percentage of them were present in Japanese bathrooms**

The Ops found in the bathroom samples are well reported (Kelley *et al*. 2013; Novak *et al*. 2020; Falkinham. 2020b). *Acinetobacter*, *Stenotrophomonas*, *Pseudomonas*, *Mycobacterium* and *Legionella* are commonly presented Ops worldwide. In this study, we encountered two pathogen-like species, *Staphylococcus aureus* (87.98% maximum value of identity) and *Vibrio parahaemolyticus* (87.86% maximum value of identity). *S. aureus* is known as a commensal bacterium and as a pathogen that causes opportunistic infections mainly of soft tissues, skin, and wounds (Kozajda *et al*. 2019). Studies have shown that the detection rate is higher in areas with high hand and foot contact (Ojima *et al*. 2002). Infection with *V. parahaemolyticus* can cause gastroenteritis, septicaemia, and infection. It is a seafood-associated and a water-borne food-mediated bacterium (Liu *et al*. 2015). This bacterium has been detected in freshwater in some cases (Maje *et al*. 2020; Silva *et al*. 2018). In addition to the presence of these bacteria, we also detected some Op genera that are closely related to *Corynebacterium* and *Mycobacterium*. *Corynebacterium*, a known human-associated bacterium, is commonly found in bathrooms, according to meta-analysis data from high-throughput amplicon analysis as well as from culture methods (Allen *et al*. 2004; Adams *et al*. 2015). In this study, *Legionella*, a problem in water systems around the world, was not detected. Some studies show that planktonic and biofilm *Legionella* concentrations are reduced by the presence of residual disinfectants within water material (Waak *et al*. 2018; Fish *et al*. 2020). The standard value of residual chlorine in water in Japan is set at 1 mg/L or less, is often 0.1 mg/L or less and is strictly controlled. The fact that *Legionella* was not detected in this study may indicate the high-performance level of the water quality management system in Japan. *Mycobacterium* was detected in 34% (11/32) of the samples, and the mean and median values were 0.40% and 0.00%, respectively (Figure 8). Arikawa *et al*. surveyed *Mycobacterium avium* subsp. *hominissuis* (MAH) in Japan, and the detection rate of MAH was 16.1% (Arikawa *et al*. 2019). Using a culture-independent method, quantitative PCR, Ichijo, *et al*. detected the frequency of *Mycobacterium* spp. in showerheads. The detection rate was lower than that in our study (13%) (Ichijo *et al*. 2014). According to Gebert *et al*., the detection rate of *Mycobacterium* spp. is 37% across households in the United States and Europe, which is similar to the results of this study. However, the relative abundances of mycobacteria were much higher than those in this study, at approximately 15% and 7%, respectively (Gebert *et al*. 2018). Additionally, Feazel *et al.* reported that *M. avium* was identified in 20% of showerhead swabs in the US (Feazel *et al*. 2009). *Mycobacterium* was not

present at a higher rate in Japan than in the US and Europe; however, the incidence rate of nontuberculous mycobacterial pulmonary disease in Japan was reported to be 14.7 cases per 100,000 people in 2014 (Namkoong *et al*. 2016), and the number of infected people is increasing rapidly compared to other countries (Yano *et al*. 2017). Non-tuberculous mycobacteria include nearly 200 species that can differ with respect to their ecology and pathogenicity (Tortoli 2014; Falkinham 2020a). Thus, to obtain more detailed information on specific mycobacterial species, it is necessary to sequence the gene using mycobacterium-specific primers, such as *hsp65* (65-kDa heat shock protein), as well as performing identification through culture methods (Telenti *et al*. 1993). In this study, we encountered the assignation of *Brucella* to other genera. Misidentification of *Brucella* species has been reported in many studies (Elsaghir and James 2003; Horvat *et al*. 2011; Carrington *et al*. 2012), except for in those using near-full-length 1,412-bp nucleotide sequences of 16S RNA genes and Sanger sequencing (Gee *et al*. 2004). Because of these recurring misidentifications, an increasing number of laboratories are now relying on molecular methods such as Gram staining, semisolid-medium motility tests and flagellar staining to identify *Brucella* (Yang *et al*. 2013). In the results of the sequence analysis in this study, even if *Brucella* was detected, it was necessary to validate the results by molecular methods using cultured strains.

One of the major advantages of using targeted metagenomic techniques, such as 16S rRNA gene sequencing, is that these techniques are culture-independent and can theoretically recover almost all bacterial taxa in any habitat. Therefore, it is possible to identify the microorganisms, including the percentage of potential pathogens, in the entire community. Moreover, MinION can generate long read length. The increased information content inherent from longer read lengths assists researchers with alignment-based taxonomy assignment (Wommack *et al*. 2008). Mitsuhashi *et al*. and Nakagawa *et al*. reported that 5-min and 3-min running times on MinION, respectively, were sufficient to detect specific bacteria (Mitsuhashi *et al*., 2017; Nakagawa *et al*., 2019). Furthermore, by optimizing the PCR conditions, MinION was shown to provide accurate microbial community composition and more accurate data than Miseq in terms of species-level matches (Fujiyoshi *et al*. 2020). However, because of the widespread use of DNA sequencing technology, it is necessary to carefully investigate the analysis methods, databases, and results of these methods before applying them. The tendency is to use top hits to classify microorganisms, but in general, cut-offs of 95% and 98.7% are used to classify bacterial isolates at the genus and species levels, respectively (Stackebrandt and Goebel 1994). This point needs to be discussed carefully, especially when mentioning Op bacteria in samples. Additionally, DNA does not provide information whether a microorganism is alive, dead, or infectious. Therefore, it is necessary and efficient to conduct epidemiological studies using both metagenomic approaches and culture methods simultaneously. In 2020, we published a paper on a "suitcase lab" in which a single suitcase can contain all the necessary equipment from sampling to detection by the LAMP method and showed that the work can be completed on-site within two hours if the specific target species is decided upon (Fujiyoshi *et al*. 2021). In the future, if such a tool can be used to analyse the microbiome on-site quickly and easily, it will contribute to the improvement of microbial risk assessment and control not only of water quality but also of the built environment.

Our comprehensive analysis of the bacterial community in the built environment revealed that specific bacteria in source water would primarily adhere to the surface of the showerhead where they subsequently form biofilms. Moreover, the bacterial communities in biofilms differed greatly among samples. This probably suggests that the bacterial community in a biofilm is influenced by factors other than the water. The findings of this study are the first data in Japan for assessing microbial communities in built environments. This is also the first study to analyse the microbial communities of both water and its biofilm obtained from same bathrooms, by which the microbial similarity and difference of water and biofilm were emphasized clearly. We suggest that it is important to manage risks of infection in each household and that this requires rapid and easy on-site identification of microbial communities.

Methods

Sampling

A total of 50 showerhead biofilms and shower water samples from 25 independent bathrooms were collected from five regions in Japan: 3 from Tohoku, 2 from Chubu, 7 from Kanto and 12 from Kinki, as shown in Figure 1B and Supplemental Table S1 (the sampling sites were classified according to Arikawa *et al*. 2019). The showerhead biofilm and shower water samples in the same bathroom were aseptically collected. Showerhead biofilms were swabbed with sterilized nylon swabs (FLOQ swab 552C; Becton, Dickinson and Company, Tokyo, Japan), and 2 L of shower water was filtered on-site using a portable peristaltic pump (Sentino microbiology pump, Pall Life Science, MI) and a 0.2-μm filter cartridge (Sterivex, Millipore, MA, USA) (Figure 1). The cartridge was put into a sterile tube, and then the samples were immediately placed in a cool box at 4°C, transported to the lab within a few hours and kept at -20°C until use.

Scanning electron microscope (SEM)

315 The biofilms on coverslips and filters were prefixed in a solution of 2.5% glutaraldehyde in 0.1
316 M phosphate buffer (PB; pH 7.4) for 10 min and rinsed three times with PB. Samples were then
317 fixed again with 2.5% glutaraldehyde for 1 h and rinsed three times with PB. Another fixation
318 reagent, 1% (w/v) osmium tetroxide in PB, was added to the samples, followed by incubation for
319 1 h. Subsequently, the samples were rinsed three times with PB and dehydrated with increasing
320 concentrations of ethanol (30%, 50%, 70%, 90%, 99%, and 100%). Dehydrated samples were
321 soaked in isoamyl acetate, successively critical-point-dried with an HCP-2 (Hitachi Ltd., Tokyo,
322 Japan), and coated with 8:2 platinum–palladium alloy using an E-1030 ion sputter (Hitachi Ltd.,
323 Tokyo, Japan). The resultant coating was 12 nm thick. Samples were observed using an S4700
324 scanning electron microscope (Hitachi Ltd.).

325

326 **DNA extraction and PCR conditions**

All samples were recovered using aseptic techniques and appropriate negative and positive controls. A swab and a filter, picked out from the cartridge, were directly placed in a bead tube of a DNeasy PowerBiofilm Kit (QIAGEN, Germantown, MD, USA) under a laminar flow cabinet, and DNA was extracted according to the manufacturer's protocol with some modifications (Arai *et al*. 2018): instead of glass beads in a PowerBiofilm bead tube, 400 µL of sterilized φ0.5 mm zirconia beads (TORAY, Tokyo, Japan) and two gains of φ5 mm zirconia beads (TORAY) were used for homogenization. The samples were bead beaten with a bead crusher (TissueLyser II, QIAGEN) at 3,200 rpm for 10 min. The DNA was eluted in 100 µL of elution buffer and then purified and condensed with a Dr. GenTLE precipitation carrier (Takara BIO, Tokyo, Japan). The concentration and purity of the DNA were measured with a DS-11FX+ Spectro/Fluorometer (DeNovix, Wilmington, USA) and a QuantiFluor™ dsDNA System (Promega, Madison, USA). PCR amplification and barcoding of 16S rRNA genes were conducted using the 16S Barcoding Kit (SQK-RAB204; Oxford Nanopore Technologies, Oxford, UK) containing the 27F/1492R primer set and MightyAmp DNA polymerase Ver.3 (Takara Bio). PCR was performed according to a previous report (Fujiyoshi *et al*. 2020). A reaction containing no template served as the negative control. The amplified fragments were separated in a 2% agarose gel, stained with Safelook Load Green (Wako Chemicals Co. Ltd, Osaka, Japan), and checked with a FAS Nano Gel Document System (NIPPON Genetics, Tokyo, Japan).

Nanopore sequencing library construction

After purifying a PCR product (50 µl) with 30 µl of Agencourt AMPure XP (Bechman Coulter, Tokyo, Japan), the amount and purity of the DNA eluted with 10 µl of buffer solution (pH 8.0, 10 mM Tris-HCl with 50 mM NaCl) were determined as above. One hundred or 50 fmol of purified amplicon DNA was used as input DNA for MinION-compatible libraries. The amplicons were added to 1 µl of rapid adapter (Oxford Nanopore Technologies) and incubated at room temperature for 5 min.

Sequencing and data analysis

Each nanopore sequencing library was run on FLO-MIN106 R9.4 flow cells (Oxford Nanopore Technologies) after performing platform QC analysis. The amplicon library (11 µl) was diluted with running buffer (35 µl), with 3.5 µl of nuclease-free water, and with 25.5 µl of loading beads. A 48-h sequencing protocol was initiated using MinION control software MinKNOW v.1.11.5 or 1.14.1. MinION sequence reads (i.e., FAST5 data) were converted into FASTQ files using Albacore v.2.1.3 or 2.3.3 software (Oxford Nanopore Technologies). FASTQ files were analysed as described previously (Fujiyoshi *et al*. 2020). The files examined the sequence read length distribution using FastQC (v 0.11. 2) (Andrews. 2010), and Seqkit 0.8.0 (https://bioinf.shenwei.me/seqkit/) (Shen *et al*. 2016) was used to filter the sequence data by the lengths of 1,400–1,600 to include 1,500–base reads. After filtering, the sequence reads were mapped using bwa-mem (v. 0.7. 17) (Li 2013), with the MinION analysis option (-x ont2d) (Jain *et al*. 2015), to a database derived from the Ribosomal Database Project (RDP Release 11, Update 5, Seqt. 30. 2016) (Cole *et al*. 2014), and the top hit was used for genus and species assignment. The RDP hierarchy browser (http://rdp.cme.msu.edu/hierarchy/hb_intro.jsp) was used with the following filters: strain = "Type"; source = "isolates"; size ">= 1,200"; quality = "Good"; and taxonomy = "Nomenclatural" (for analytical scheme, see Figure 1C). Sequences from all the samples were normalized to the sample containing the lowest number of reads (5115 reads). After the removal of singleton OTUs, all data analysis was carried out with R (v. 3.3.1) (R Core Team 2018). The R package vegan (v. 2.5-5) was applied for diversity and community analyses. Two metrics of alpha diversity were used in this study, Shannon diversity index (richness and evenness) and Chao1 index (richness). Significant differences in the number of OTUs and alpha diversity metrics between samples were determined using paired t-tests. Beta diversity was explored by non-metric multidimensional scaling (nMDS) of Bray-Curtis dissimilarity among samples. Statistical significance was calculated by permutational analysis of variance (PERMANOVA). Venn diagrams were visualized with the R package 'VennDiagram' (Chen and Boutros 2011), illustrating the shared and unique genera among each bathroom paired sample. Linear discriminant analysis (LDA) effect size (LEfSe) was applied to identify specific bacterial genera between the samples (Segata *et al*. 2011). Taxa were considered significant based on LDA scores greater than 4.0 and p-values smaller than 0.05.

Possibility of pathogens in the bathroom

The presence or absence of genera on the National Institute of Infectious Diseases (NIID) in Japan list of Bio Safety Level (BSL) 2 and 3 were checked in each sample (for the genus in NIID list, see Supplemental Table S4, and for the list of species, see the following link: https://www.nite.go.jp/nbrc/mrinda/list/risk/bacteria/ALL, updated 2010-06). As a result, each read corresponding to BSL2 and BSL3 species was collected, and then sequences longer than 1,000 bp were filtered by seqkit (v0.8.0). The sequences were clustered with psi-cd-hit.pl to identify representative sequences (Li and Godzik 2006; Fu *et al*. 2012). The settings in psi-cd-hit.pl were the default except for -prog blastn. For further analysis, the sequences with the highest number of reads in each cluster were used. Taxonomy was identified with BLASTn (E value = 1.0e-10, word size = 7, reward = 2, penalty = -3, gap open = 5, gap extend = 2, filter: unmarked "low complexity regions") (Camacho *et al*. 2009) against 16S ribosomal RNA from curated type strain sequences from bacteria and archaea in GenBank (update date: 2021/01/11, number of sequences: 21,699). The analysis scheme is shown in Figure 1C.

Data availability

The authors declare that all the data supporting the findings of this study are available within the paper (and its Supplemental files) and that raw data were presented where possible. The raw MiSeq data reported in the paper (Figure 6, Supplemental Figures and Tables) have been uploaded to the DDBJ database under the accession number DRA010182. It will be open to public when this manuscript is accepted for publication

Author contributions

S.F. and **F.M.** conceived and designed the experiment. **S.F.** and **F.M.** collected samples. **S.F.** performed the experiments and analysed the data. **S.F.**, **Y.N.**, and **F.M.** interpreted the data. **S.F.** wrote the manuscript. **S.F.**, **Y.N.**, and **F.M.** reviewed drafts of the manuscript.

Acknowledgements

This work was supported by The Kyoto University Research Funds for Young Scientists and the Japan Society for the Promotion of Science under Grants-in-Aid for Scientific Research (KAKENHI) (grant number 20K18903 to S.F.). KAKENHI (grant numbers 18K19674/18KK0436/20H00562) and the Japan Agency for Medical Research and Development (grant number 20wm0225012h0001/21fk0108129h0502) awarded grants to F.M. This manuscript was edited for English language by American Journal Experts (AJE).

Competing interests

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

References

Adams, R. I., Bateman, A. C., Bik, H. M., Meadow, J. F., 2015. Microbiota of the indoor environment: a meta-analysis. Microbiome, 3, 49. https://doi.org/10.1186/s40168-015-0108-3.

Allen, M. J., Edberg, S. C., Reasoner, D. J., 2004. Heterotrophic plate count bacteria —what is the significance in drinking water?. Int J Food Microbiol, 1, 265–74. https://doi.org/10.1016/j.ijfoodmicro.2003.08.017.

Andrews, S., 2010. Fastqc: A quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Arai, S., Kim, H., Watanabe, T., Tohya, M., Suzuki, E., Ishida-Kuroki, K., Maruyama, F., Murase, K., Nakagawa, I., Sekizaki, T., 2018. Assessment of pig saliva as a *Streptococcus suis* reservoir and potential source of infection on farms by use of a novel quantitative polymerase chain reaction assay. Am J Vet Res, 79, 941–948. https://doi.org/10.2460/ajvr.79.9.941.

Arikawa, K., Ichijo, T., Nakajima, S., Nishiuchi, Y., Yano H., Tamaru, A., Yoshida, S., Maruyama, F., Ota, A., Nasu, M., Starkova, D. A., Mokrousov, I., Narvskaya, O. V., Iwamoto, T., 2019. Genetic relatedness of *Mycobacterium avium* subsp. *hominissuis* isolates from bathrooms of healthy volunteers, rivers, and soils in Japan with human clinical isolates from different geographical areas. Infect Genet Evol, 74, 103923. https://doi.org/10.1061/j.meegid.2019.103923.

441 Boe-Hansen, R., Albrechtsen, H. J., Arvin, E., Jørgensen, C., 2002. Bulk water phase and biofilm
442 growth in drinking water at low nutrient conditions. Water Res, 36, 4477–86.
443 https://doi.org/10.1016/s0043-1354(02)00191-4.

444 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L.,
445 2009. Blast+: architecture and applications. BMC Bioinformatics, 10, 421.
446 https://doi.org/10.1186/1471-2105-10-421.

447 Cervero-Aragó, S., Rodríguez-Martínez, S., Puertas-Bennasar, A., Araujo, R. M., 2015. Effect of
448 common drinking water disinfectants, chlorine and heat, on free *Legionella* and amoebae-
449 associated *Legionella*. PLos One, 10, e0134726. https://doi.org/10.1371/journal.pone.0134726.

450 Chen, H., Boutros, P. C., 2011. Venndiagram: a package for the generation of highly-
451 customizable Venn and Euler diagrams in R. BMC Bioinformatics, 12, 35.
452 https://doi.org/10.1186/1471-2105-12-35.

453 Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-
454 Alfaro, A., Kuske, C. R., Tiedje, J. M., 2014. "Ribosomal Database Project: data and tools for
455 high throughput rRna analysis. Nucleic Acids Res, 42 (Database issue): D633–42.
456 https://doi.org/10.1093/nar/gkt1244.

457 Dannemiller, K. C., Gent, J. F., Leaderer, B. P., Peccia, J., 2016. Indoor microbial communities:
458 Influence on asthma severity in atopic and nonatopic children. J. Allergy Clin Immunol, 138, 76–
459 83.e1. https://doi.org/10.1016/j.jaci.2015.11.027.

460 Denis, C., Irlinger, F., 2007. Safety assessment of dairy microorganisms: aerobic coryneform
461 bacteria isolated from the surface of smear-ripened cheeses. Int J Food Microbiol 126: 311–5.
462 https://doi.org/10.1016/j.ijfoodmicro.2007.08018.

463 Douterelo, I., Fish, K. E., Boxall, J. B., 2018. Succession of bacterial and fungal communities
464 within biofilms of a chlorinated drinking water distribution system. Water Res, 15, 74–85.
465 https://doi.org/10.1016/j.watres.2018.04.058.

466 Elsaghir, A. A. F., James, E. A., 2003. Misidentification of *Brucella melitensis* as *Ochrobactrum*
467 *anthropi* by Api 20NE. J Med Microbiol, 52, 441–2. https://doi.org/10.1099/jmm.0.05153-0.

468 Falkinham, J. O., III, 2020a. *Mycobacterium avium* Complex (MAC) in water distribution
469 systems and household plumbing in the United States. Water, 12, 3338.
470 https://doi.org/10.3390/w12123338.

471 Falkinham, J. O., III, 2020b. Living with *Legionella* and other waterborne pathogens.
472 Microorganisms, 8, 2026. https://doi.org/10.3390/microorganisms8122026.

473 Feazel, L. M., Baumgartner, L. K., Peterson, K. L., Frank, D. N., Harris, J. K., Pace, N. R., 2009.
474 Opportunistic pathogens enriched in showerhead biofilms. Proc Natl Acad Sci U S A, 106,
475 16393–9. https://doi.org/10.1073/pnas.0908446106.

476 Fish, K. E., Reeves-McLaren, N., Husband, S., Boxall, J., 2020. Unchartered waters: the
477 unintended impacts of residual chlorine on water quality and biofilms. npj Biofilms and
478 Mirobiomes. 6, 34. https://doi.org/10.1038/s41522-020-00144.

479 Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. Cd-Hit: accelerated for clustering the next-
480 generation sequencing data. Bioinformatics, 28, 3150–2.
481 https://doi.org/10.1093/bioinformatics/bts565.

482 Fujiyoshi, S., Muto-Fujita A., Maruyama, F., 2020a. Evaluation of PCR conditions for
483 characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore
484 sequencer. Sci Rep, 10, 12580. https://doi.org/10.1038/s41598-020-69450-9.

485 Fujiyoshi, S., Yarimizu, K., Miyashita, Y., Rilling, J., Acuña, J. J., Ueki, S., Gajardo, G.,
486 Espinoza-Gonzáles, O., Guzmán, L., Jorquera, M. A., Nagai, S., Maruyama, F., 2021. Suitcase
487 Lab: new, portable, and deployable equipment for rapid detection of specific harmful algae in
488 Chilean coastal waters. Environ Sci Pollut Res Int, 28, 14144–55.
489 https://doi.org/10.1007/s11356-020-11567-5.

490 Gallego, V., García, M. T., Ventosa A., 2005. *Methylobacterium isbiliense* sp. nov., isolated
491 from the drinking water system of Sevilla, Spain. Int J Syst Evol Microbiol, 55, 2333–7.
492 https://doi.org/10.1099/ijs.0.63773-0.

493 Gebert, M. J., Delgado-Baquerizo, M., Oliverio, A. M., Webster, T. M., Nichols, L. M., Honda,
494 J. R., Chan, E. D., Adjemian, J., Dunn, R. R., Fierer, N., 2018. Ecological analyses of
495 Mycobacteria in showerhead biofilms and their relevance to human health. mBio, 9, e01614-18.
496 https://doi.org/10.1128/mBio.01614-18.

497 Gee, J. E., De, B. K., Levett, P. N., Whitney, A. M., Novak, R. T., Popovic T., 2004. Use of 16S
498 rRNA gene sequencing for rapid confirmatory identification of *Brucella* isolates. J Clin
499 Microbiol, 42, 3649–54. https://doi.org/10.1128/JCM.42.8.3649-3654.2004.

500 Horvat, R. T., El Atrouni, W., Hammoud, K., Hawkinson, D., Cowden, S., 2011. Ribosomal
501 RNA sequence analysis of *Brucella* infection misidentified as *Ochrobactrum anthropi* infection.
502 J Clin Microbiol, 49, 1165–8. https://doi.org/10.1128/JCM.01131-10.

503 Huang, G., Ng, T. W., Chen, H., Chow, A.T., Liu, S., Wong, P. K., 2020. Formation of
504 assimilable organic carbon (AOC) during drinking water disinfection: A microbiological
505 prospect of disinfection byproducts. Environ Int, 135, 105389.
506 https://doi.org/10.1016/j.envint.2019.105389.

507 Ichijo, T., Izumi, Y., Nakamoto, S., Yamaguchi, N., Nasu, M., 2014. Distribution and respiratory
508 activity of Mycobacteria in household water system of healthy volunteers in Japan. PLos One, 9,
509 e110554. https://doi.org/10.1371/journal.pone.0110554.

510 Iwamoto, T., Nakajima, C., Nishiuchi, Y., Kato, T., Yoshida, S., Nakanishi, N., Tamaru, A.,
511 Tamura, Y., Suzuki, Y., Nasu, M., 2012. Genetic diversity of *Mycobacterium avium* subsp.
512 *hominissuis* strains isolated from humans, pigs, and human living environment. Infec Genet
513 Evol, 12, 846–52. https://doi.org/10.1061/j.meegid.2011.06.018.

514 Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., Akeson, M., 2015. Improved data
515 analysis for the MinION nanopore sequencer. Nat Methods, 12, 351–6.
516 https://doi.org/10.1038/nmeth.3290.

517 Ji, P., Rhoads W. J., Edwards, M. A., Pruden, A., 2017. Impact of water heater temperature
518 setting and water use frequency on the building plumbing microbiome. ISME J, 11, 1318–30.
519 https://doi.org/10.1038/ismej.2017.14.

520 Jing, Z., Lu, Z., Mao, T., Cao, W., Wang, W., Ke, Y., Zhao, Z., Wang, X., Sun, X. 2021.
521 Microbial composition and diversity of drinking water: A full scale special-temporal
522 investigation of a city in northern China. Sci Total Environ, 776, 145986.
523 https://doi.org/10.1016/j.scitotenv.2021.145986.

524 Johnson L. R., 2008. Microcolony and biofilm formation as a survival strategy for bacteria. J
525 Theor Biol, 7, 24–34. https:doi.org/j.jtbi.2007.10.039.

526 Kawai, M., Ichijo, T., Takahashi, Y., Noguchi, M., Katayama, H., Cho, O., Sugita, T., Nasu, M.,
527 2019. Culture independent approach reveals domination of human-oriented microbes in a
528 pharmaceutical manufacturing facility. Eur J Pharm Sci, 137, 104973.
529 https://doi.org/j.ejps.2019.104973.

530 Kelley, S.T., Gilbert, J.A., 2013. Studying the microbiology of the indoor environment. Genome
531 Biol, 14, 202. https://doi.org/10.1186/gb-2013-14-2-202.

532 Kelley, S. T., Theisen, U., Angenent, L. T., St Amand, A., Pace, N. R., 2004. Molecular analysis
533 of shower curtain biofilm microbes. Appl Environ Microbiol, 70, 4187–92.
534 https://doi.org/10.1128/AEM.70.7.4187-4192.2004.

535 Kozajda, A., Jeżak, K., Kapsa, A., 2019. Airborne *Staphylococcus aureus* in different
536 environments—a review. Environ SciPollut Res Int, 26, 34741–53.
537 https://doi.org/10.1007/s11356-019-06557-1.

538 Kormas, K. A., Neofitou, C., Pachiadaki, M., Koufostathi, E., 2010. Changes of bacterial
539 assemblages throughout an urban drinking water distribution system. Environ Monit Assess, 165,
540 27–38. https://doi.org/10.1007/s10661-009-0924-7.

541 Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
542 arXiv preprint, arXiv, 1303.3997.

543 Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of
544 protein or nucleotide sequences. Bioinformatics, 22, 1658–9.
545 https://doi.org/10.1093/bioinformatics/btl158.

546 Li, W., Tan, Q., Zhou, W., Chen, J., Li, Y., Wang, F., Zang, J., 2020. Impact of substrate
547 material and chlorine/chloramine on the composition and function of a young biofilm microbial
548 community as revealed by high-throughput 16S rRNA sequencing. Chemosphere, 242, 125310.
549 https://doi.org/10.1016/j.chemosphere.2019.125310.

550 Liu, Y., Hung, Y. H., Yuan, J., Chen, F,m Cai, W., Liu, J., Ma, X., Xie, C., Zheng, C., Zhuo, L.,
551 Cao, X., Tan, H., Li, B., Xie, H., Liu, Y., Ip, D., 2015. A foodborne outbreak of gastroenteritis
552 caused by *Vibrio parahaemolyticus* and Norovirus through non-seafood vehicle. PLos One, 10,
553 e0137848. https://doi.org/10.1371/journal.pone.0137848.

554 Ma, X., Li, G., Chen, R., Yu, Y., Tao, H., Zhang G., Shi B., 2020. Revealing the changes of
555 bacterial community from water source to consumers tap: A full-scale investigation in eastern
556 city of China. J Environ Sci, 87: 331–40. https://doi.org/10.1016/j.jes.2019.07.017.

557 Maje, M. D., Kaptchouang Tchatchouang, C. D., Manganyi, M. C., Fri, J., Ateba, C. N., 2020.
558 Characterisation of *Vibrio* species from surface and drinking water sources and assessment of
559 biocontrol potentials of their bacteriophages. Int J Microbiol, 2020, 8863370.
560 https://doi.org/10.1155/2020/8863370.

561 Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., Imanishi,
562 T., 2017. A portable system for rapid bacterial composition analysis using a nanopore-based
563 sequencer and laptop computer. Sci Rep, 7, 5657. https://doi.org/10.1038/s41598-017-05772-5.

564 Montagna, M. T., Cristina, M. L., De Giglio, O., Spagnolo, A. M., Napoli, C., Cannova, L.,
565 Deriu, M. G., Delia, S. A., Giuliano, A., Guida, M., Laganà, P., Liguori, G., Mura, I., Pennino,
566 F., Rossini, A., Tardivo, S., Torre, I., Torregrossa, M. V., Villafrate, M. R., Albertini, R.,
567 Pasquarella, C., 2016. Serological and molecular identification of *Legionellaw* spp. isolated from
568 water and surrounding air samples in Italian healthcare facilities. Environ Res, 146, 47–50.
569 https://doi.org/10.1016/j.envres.2015.12.015.

570 Nakagawa, S., Inoue, S., Kryukov, K., Yamagishi, J., Ohno, A., Hayashida, K., Nakazwe, R.,
571 Kalumbi, M., Mwenya, D., Asami, N., Sugimoto, C., Mutengo, M. M., Imanishi, T., 2019. Rapid
572 sequencing-based diagnosis of infectious bacterial species from meningitis patients in Zambia.
573 Clin Transl Immunology, 8, e01087. https://doi.org/10.1002/cti2.1087.

574 Namkoong, H., Kurashima, A., Morimoto, K., Hoshino, Y., Hasegawa, N., Ato, M., Mitarai, S.,
575 2016. Epidemiology of pulmonary nontuberculous mycobacterial disease, Japan. Emerg Infect
576 Dis, 22, 1116–7. https://doi.org/10.3201/eid2206.151086.

577 Neu, L., Bänziger, C., Proctor, C. R., Zhang, Y., Liu W. T., Hammers, F., 2018. Ugly
578 ducklings—the dark side of plastic materials in contact with potable water. npj Biofilms
579 Microbiomes, 4, 7. https://doi.org/10.1038/s41522-018-0050-9.

580 Nishiuchi, Y., Tamura, A., Kitada, S., Taguri, T., Matsumoto, S., Tateishi, Y., Yoshimura, M.,
581 Ozeki, Y., Matsumura, N., Ogura, H., Maekura, R., 2009. *Mycobacterium avium* complex
582 organisms predominantly colonize in the bathtub inlets of patients' bathrooms. Jpn J Infect Dis,
583 62, 182–6.

584 Nishiuchi, Y., Maekura, R., Kitada, S., Tamaru, A., Taguri, T., Kira, Y., Hiraga, T., Hirotani, A.,
585 Yoshimura, K., Miki, M., Ito, M., 2007. The recovery of *Mycobacterium avium-intracellulare*
586 complex (MAC) from the residential bathrooms of patients with pulmonary MAC. Clin Infect
587 Dis, 45, 347–51. https://doi.org/10.1086/519383.

588 Nishiuchi, Y., Iwamoto, T., Maruyama, F., 2017. Infection sources of a common non-
589 tuberculous mycobacterial pathogen, *Mycobacterium avium* complex. Front Med, 4, 27.
590 https://doi.org/10.3389/fmed.2017.00027.

591 Novak Babič, M. Gostinčar, C., Gunde-Cimerman, N., 2020. Microorganisms populating the
592 water-related indoor biome. Appl Microbiol Biotechnol, 104, 6443–62.
593 https://doi.org/10.1007/s00253-020-10719-4.

594 Odamaki, T., Bottacini, F., Mitsuyama, E., Yoshida, K., Kato, K., Xiao, J. Z., van Sinderen, D.,
595 2019. Impact of a bathing tradition on shared gut microbes among Japanese families. Sci Rep, 9,
596 4380. https://doi.org/10.1038/s41598-019-40938-3.

597 Ojima, M., Toshima, Y., Koya, E., Ara, K., Tokuda, H., Kawai, S., Kasuga, F., Ueda, N., 2002.
598 Hygiene measures considering actual distributions of microorganisms in Japanese households. J
599 Appl Microbiol, 93, 800–9. https://doi.org/10.1046/j.1365-2672.2002.01746.x.

600 Parker, J., Helmstetter, A. J., Devey, D., Wilkinson, T., Papadopulos, A. S. T., 2017. Field-based
601 species identification of closely-related plants using real-time nanopore sequencing. Sci Rep, 7,
602 8345. https://doi.org/10.1038/s41598-017-08461-5.

603 Peng, H., Zhang, Y., Wang, R., Liu, J., Liu, W. T., 2020. Assessing the contribution of biofilm to
604 bacterial growth during stagnation in shower hoses. Water Supply, 20, 2564–76.
605 https://doi.org/10.2166/ws.2020.161.

606 Proctor, C. R., Reimann, M., Vriens, B., Hammes, F., 2018. Biofilms in shower hoses. Water
607 Res, 131, 274–86. https://doi.org/10.1016/j.watres.2017.12.027.

608 R Core Team, 2018. R: a language and environment for statistical computing. R Foundation for
609 Statistical Computing, Vienna, Austria. http://www.R-project.org.

610 Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., Khan, A.
611 M., Woodford, N., Saunders, N. J., Wain, J., O'Grady, J., Livermore, D. M., 2016. Identification
612 of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-
613 based metagenomic sequencing. J Antimicrob Chemother, 72, 104–14.
614 https://doi.org/10.1093/jac/dkw397.

615 Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., Huttenhower, C.,
616 2011. Metagenomic biomarker discovery and explanation. Genome Biol, 12, R60.
617 https://doi.org/10.1186/gb-2011-12-6-r60.

618 Shen, W., Le, S., Li, Y., Hu, F., 2016. Seqkit: a cross-platform and ultrafast toolkit for FASTA/Q
619 file manipulation. PLos One, 11, e0163962. https://doi.org/10.1371/journal.pone.0163962.

620 Silva, I. P., Carneiro, C. S., Saraiva, M. A. F., Oliveira, T. A. S., Sousa, O. V., Evangelista-
621 Barreto, N. S., 2018. Antimicrobial resistance and potential virulence of *Vibrio*
622 *parahaemolyticus* isolated from water and bivalve mollusks from Bahia, Brazil. Mar Pollut Bull,
623 131, 757–62. https://doi.org/10.1016/j.marpolbul.2018.05.007.

624 Stackebrandt, E., Goebel, B. M., 1994. Taxonomic note: a place for DNA-DNA reassociation
625 and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol
626 Microbiol, 44, 846–9. https://doi.org/10.1099/00207713-44-4-846.

627 Telenti, A, Marchesi, F., Balz, M., Bally, F., Böttger, E. C., Bodmer, T., 1993. Rapid
628 identification of mycobacteria to the species level by polymerase chain reaction and restriction
629 enzyme analysis. J Clin Microbiol, 31, 175–8. https://doi.org/10.1128/JCM.31.2.175-178.1993.

630 Tochihara, Y., 1999. Bathing in Japan: A Review. J Human Environment System, 3, 27–34.

631 Tortoli, E., 2014. Microbiological features and clinical relevance of new species of the genus
632 *Mycobacterium*. Clin Microbiol Rev, 27, 727–52. http://doi.org/10.1128/CMR.00035-14.

633 Waak, M., Lapara, T. M., Halle, C., Hozalski, R. M., 2018. Occurrence of Legionella spp. In
634 water-main biofilms from two drinking water distribution systems. Environ. Sci. Technol, 52,
635 7630–9. https://doi.org/10.1021/acs.est.8b01170.

636 White, D. C., Sutton, S. D., Ringelberg, D. B., 1996. The genus *Sphingomonas*: physiology and
637 ecology. Curr Opin Biotechnol, 7, 301–6. https://doi.org/10.1016/s0958-1669(96)80034-6.

638 Wommack, K. E., Bhavsar, J., Ravel, J., 2008. Metagenomics: read length matters. Appl Environ
639 Microbiol, 74, 1453–63. https://doi.org/10.1128/AEM.02181-07.

640 Yang, J., Ren, X. Q., Chu, M. L., Meng, D. Y., Xue, W. C., 2013. Mistaken identity of *Brucella*
641 infection. J Clin Microbiol, 51, 2011. https://doi.org/10.1128/JCM.03222-12.

642 Yano, H., Iwamoto,T., Nishiuchi, Y., Nakajima, C., Starkova, D. A., Mokrousov, I., Narvskaya,
643 O., Yoshida, S., Arikawa, K., Nakanishi, N., Osaki, K., Nakagawa, I., Ato, M., Suzuki, Y.,
644 Maruyama, F., 2017. Population structure and local adaptation of MAC lung disease agent
645 *Mycobacterium avium* subsp. *hominissuis*. Genome Biol Evol, 9 2403–17.
646 https://doi.org/10.1093/gbe/evx183.

647 Yano, T., Kubota, H., Hanai, J., Hitomi, J., Tokuda, H., 2013. Stress tolerance of
648 Methylobacterium biofilms in bathrooms. Microbes Environ, 28, 87–95.
649 https://doi.org/10.1264/jsme2.me12146.

650 Zhou, Z., Xu, L., Zhu, L., Liu, Y., Shuai, X., Lin, Z., Chen, H., 2021. Metagenomic analysis of
651 microbiota and antibiotic resistome in household activated carbon drinking water purifiers.
652 Environ Int, 148, 106394. https://doi.org/10.1016/j.envint.2021.106394.
653

Figure legends

Figure 1. Sampling and sequencing procedure of showerhead biofilms and shower water. Sampling place and method (A). Out of the 25 samples collected from five regions in Japan, 16 samples were analysed. The denominator indicates the number of samples collected, and the numerator indicates the number of samples available for analysis (B). The analysis scheme is divided into two parts: community (left) and pathogen analysis (right) (C).

Figure 2. SEM image of a biofilm on the inner surface of a showerhead obtained from one water distributor.

Figure 3. Alpha-diversity measured using the Shannon index. Each box plot represents the diversity distribution of samples from water and biofilm; ** $p < 0.01$.

Figure 4. Non-metric multidimensional scaling (nMDS) analysis based on Bray-Curtis dissimilarity metrics of samples labelled according to the sample sources. A clear separation between biofilm bacterial communities and water communities was revealed (PERMANOVA, F-value=3.416, R-squared=0.096, p-value=< 0.01). Ellipses represent the 95% confidence interval. Pink: Biofilm, Blue: water

Figure 5. Comparison of the shared genera in biofilm and water samples from the same bathroom. Venn diagrams show a comparison of the percentage of shared genera for biofilms (pink) and water (blue) collected from the same bathroom. Parentheses indicate the percentage of the shared genus sequences in each sample.

Figure 6. Taxonomic composition of the bacterial community in the biofilm and water samples at the genus level using a stacked bar plot (top 10 genera).

Figure 7. LEfSe analysis of the bacterial community of water and biofilm. Both water and biofilm samples were analysed, but no indicator genera were detected in biofilm. Features are significant based on their adjusted p-value and LDA score. The adjusted p-value cut-off=0.05 and LDA score=4.0.

Figure 8. This heatmap table summarizes the results for all showerhead swab and shower water libraries pooled at the genus level and grouped by region in Japan. Figure footnotes: – signifies that the genera were not detected in the sample; SH indicates a biofilm sample and W indicates a water sample.

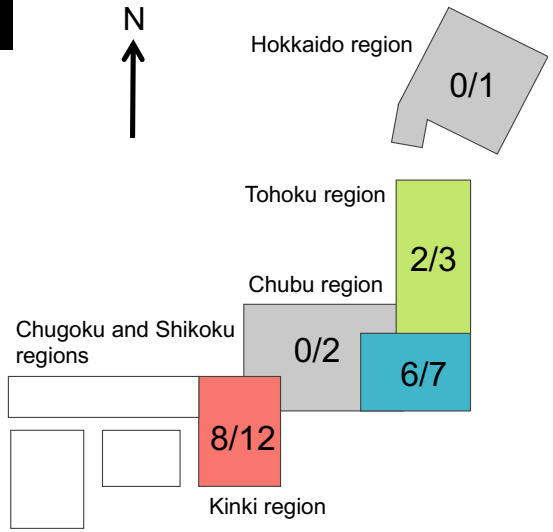
Supplemental Figure S1. Taxonomic composition of the bacterial community in each sample at the genus level using a stacked bar plot (top 30 genera).

A

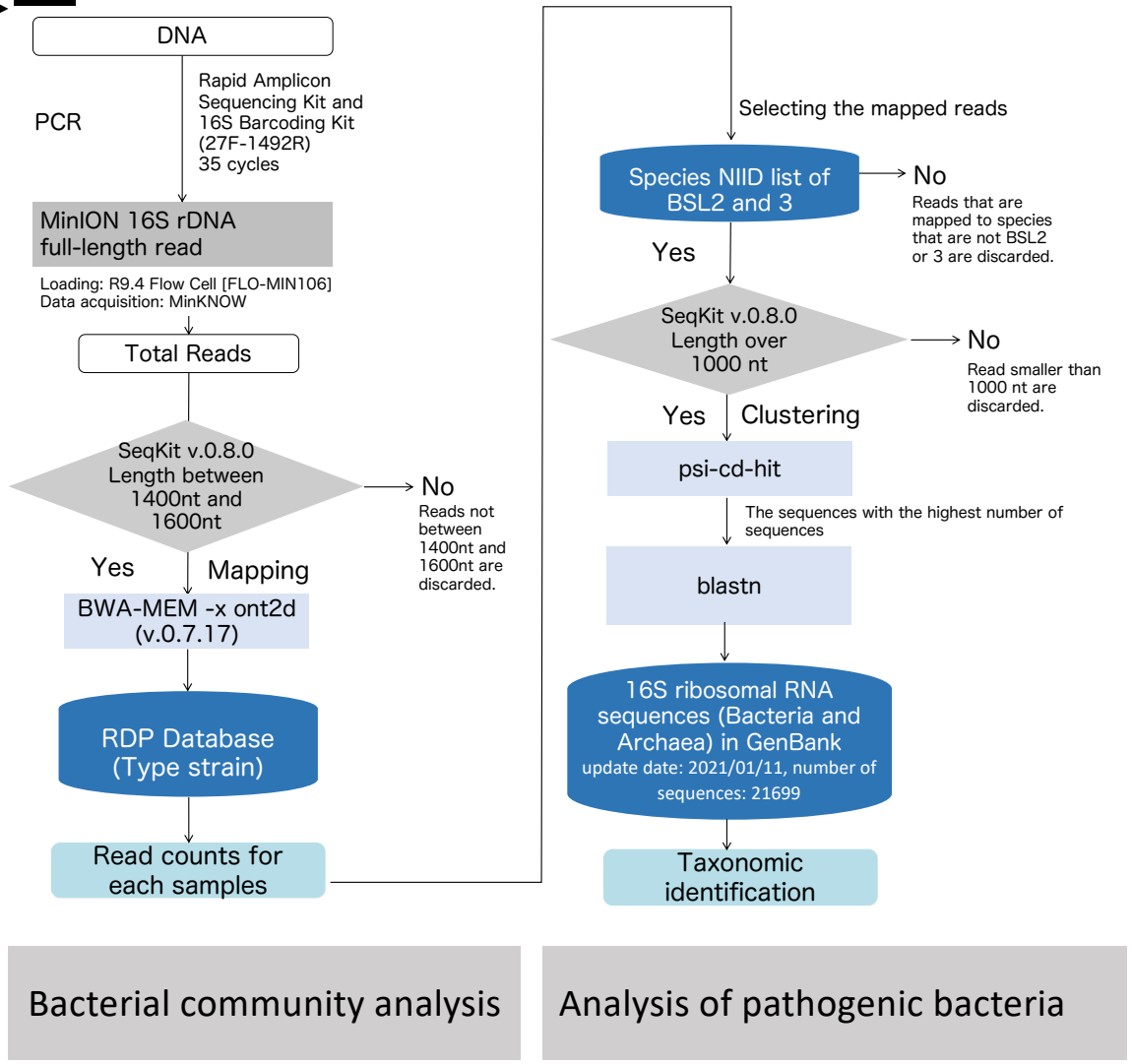
1. Shower water
2. Collection
3. Filtration

DNA extraction

B

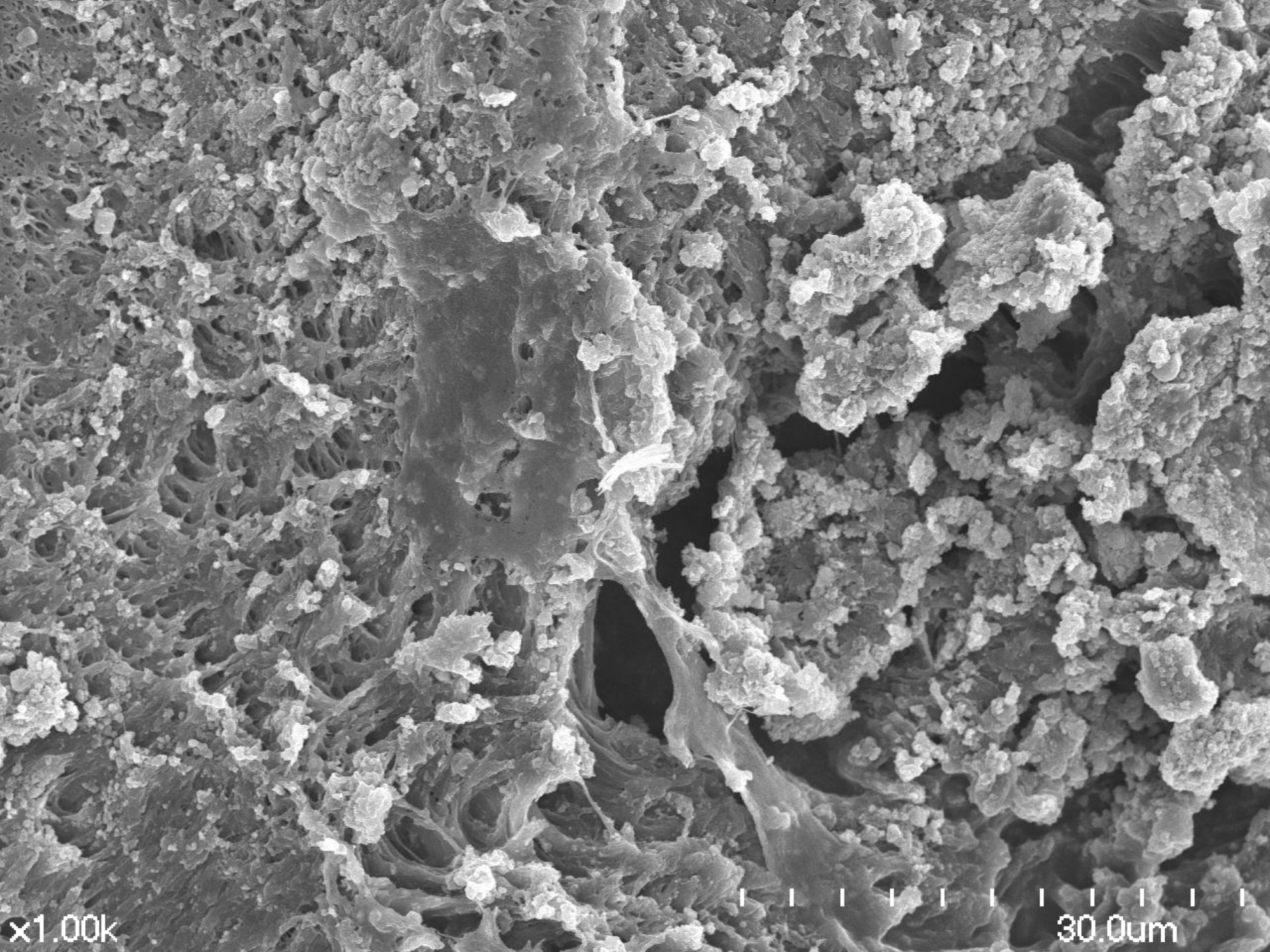


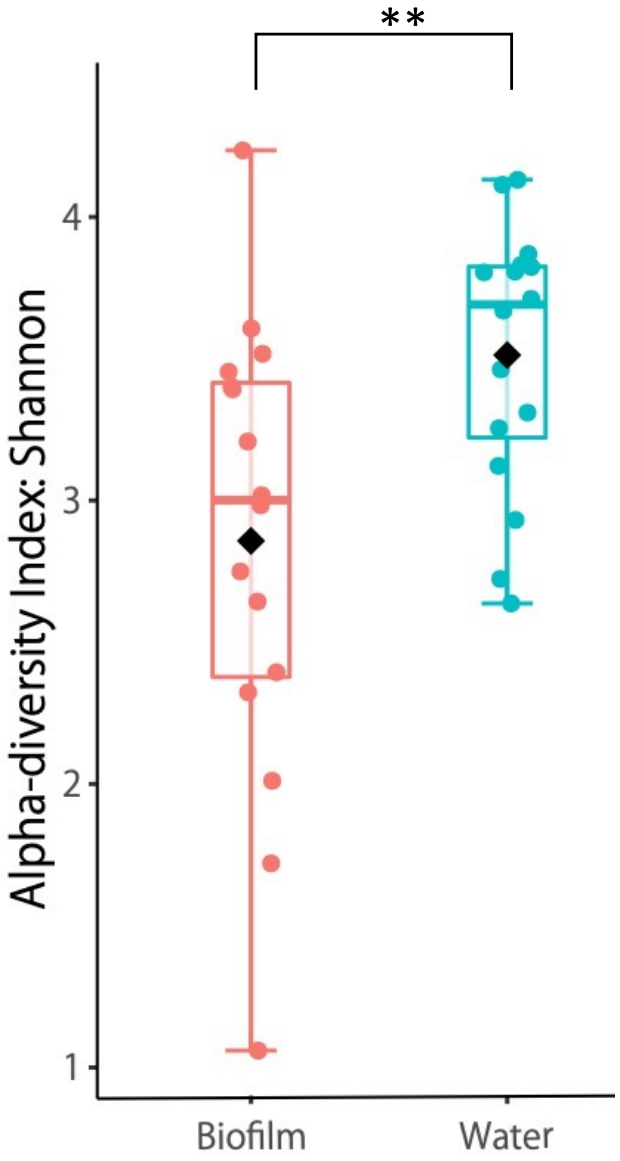
C

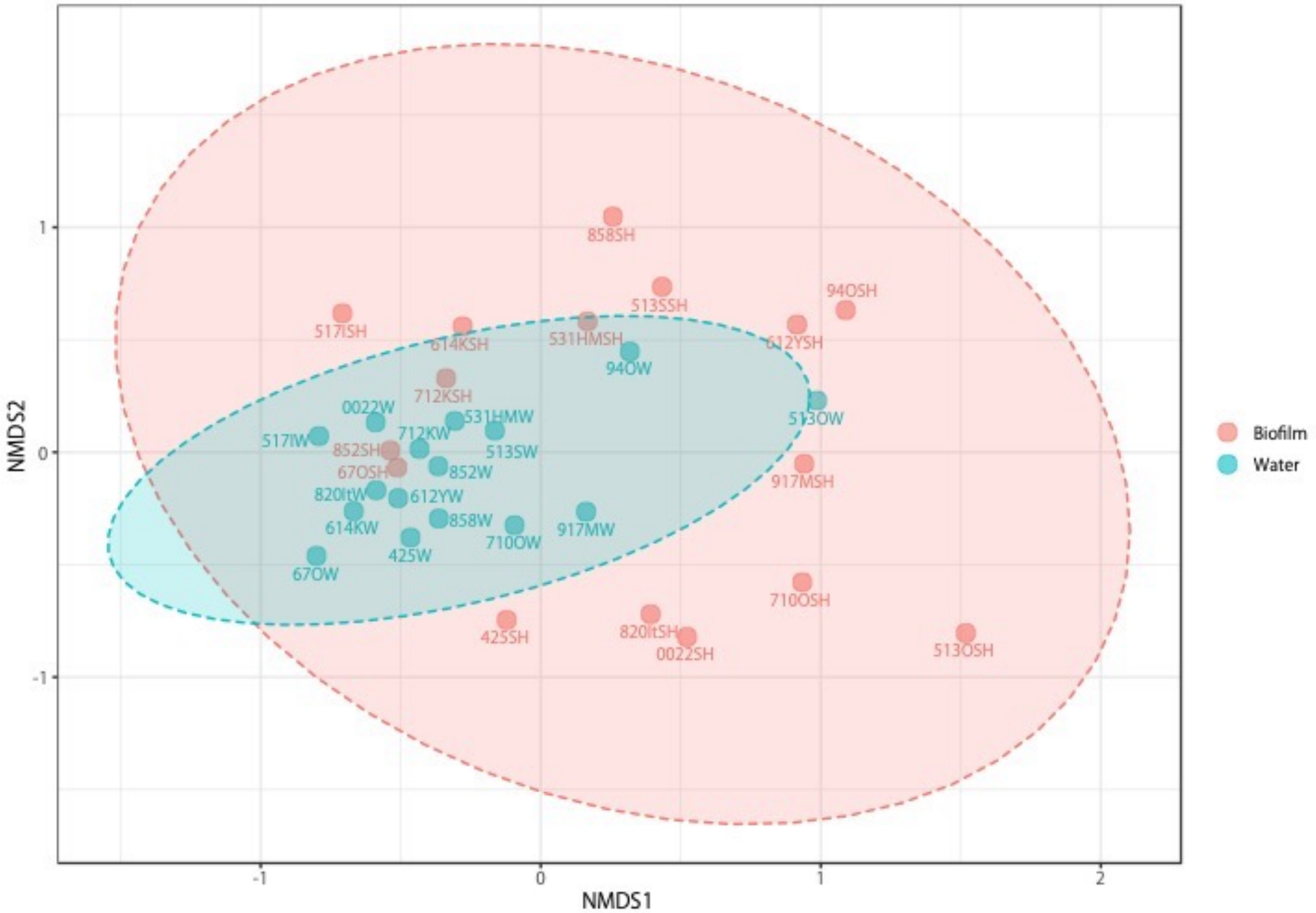


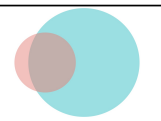
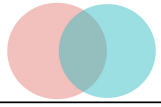
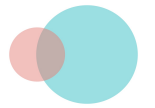
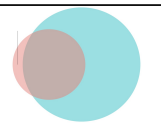
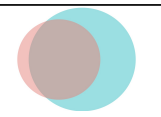
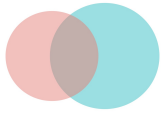
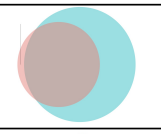
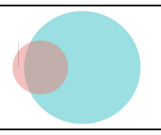
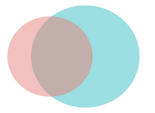
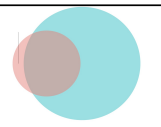
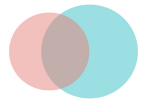
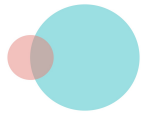
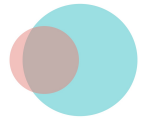
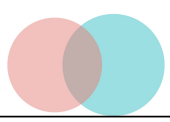
Bacterial community analysis

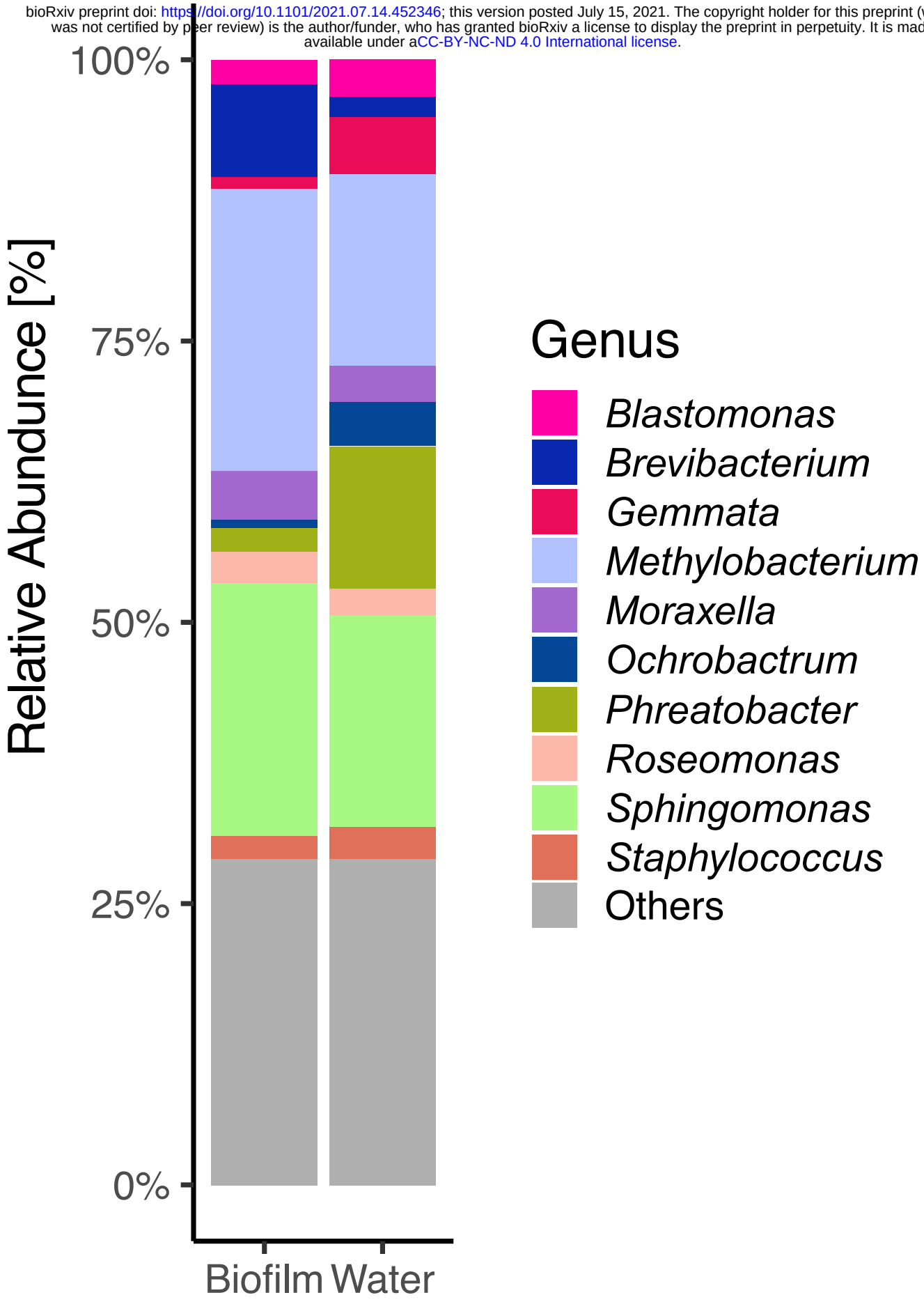
Analysis of pathogenic bacteria

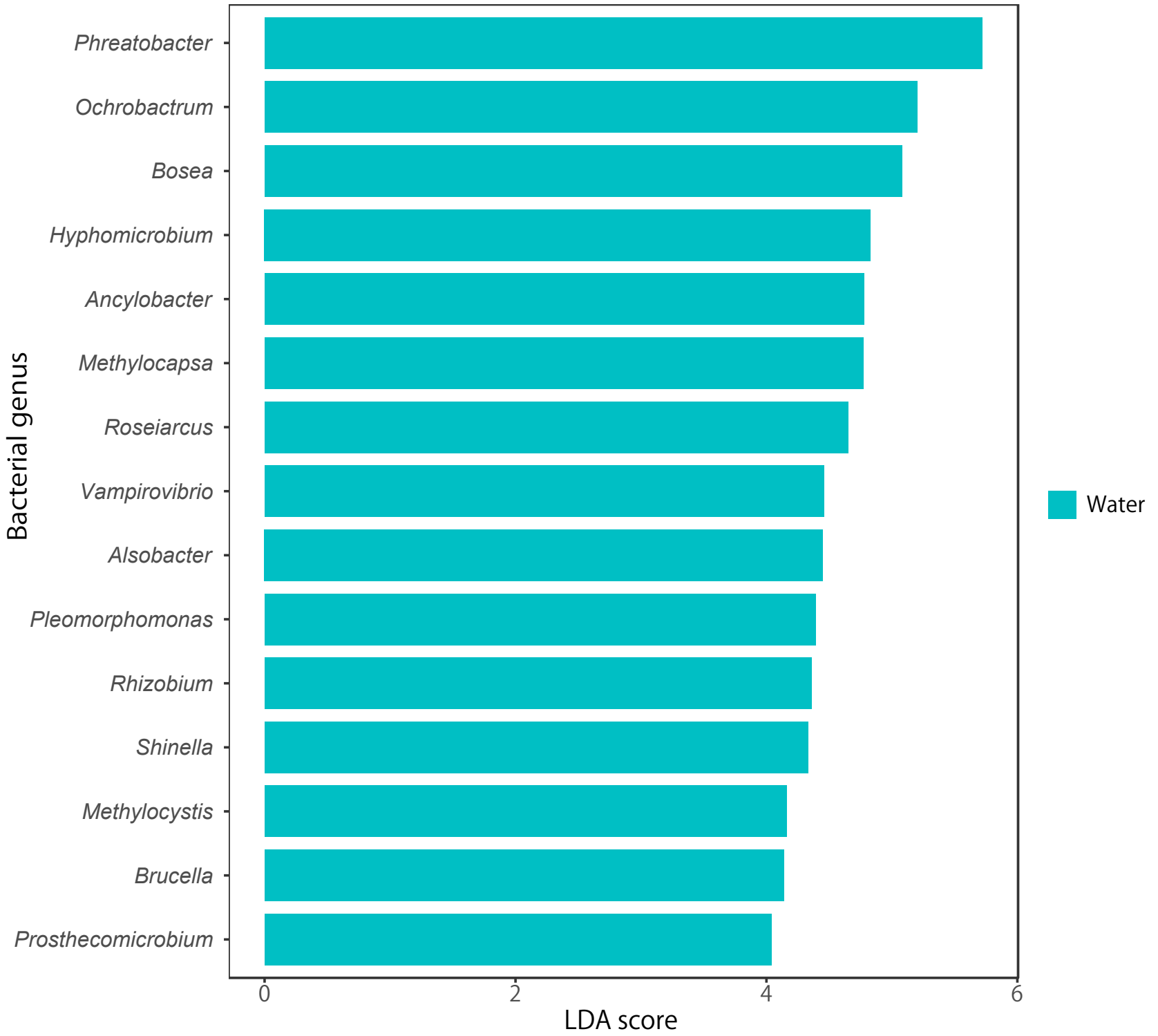

x1.00k
30.0um


**
Alpha-diversity Index: Shannon
4
3
2
1
Biofilm
Water


858SH
513SSH
940SH
517ISH
614KSH
531HMSH
612YSH
940W
712KSH
513OW
0022W
531HMW
517IW
712KW
513SW
852SH
852W
67OSH
917MSH
820ItW
612YW
858W
614KW
917MW
710OW
425W
67OW
710OSH
425SH
820ItSH
0022SH
513OSH
Biofilm
Water
NMDS1
NMDS2
-1
0
1
2
-1
0
1

Sample name	Biofilm [%]	Shared genus /Total genus	Water [%]
423SH-423W	98.7 (84.9)		62.9 (60.8)
858SH-858W	78.6 (99.8)		23.9 (50.5)
513OSH-513OW	36.8 (45.3)		38.9 (89.6)
94OSH-94OW	46.2 (98.4)		14.0 (76.3)
67OSH-67OW	76.3 (96.2)		77.6 (98.6)
710OSH-710OW	88.2 (99.8)		33.7 (72.5)
712KSH-712KW	86.8 (98.8)		52.3 (89.8)
820ItSH-820ItW	43.5 (25.2)		31.3 (50.9)
852SH-852W	93.6 (99.9)		51.2 (87.5)
513SSH-513SW	81.3 (99.9)		18.6 (79.2)
517ISH-517IW	68.9 (97.5)		42.5 (62.0)
531HMSH-531HMW	85.7 (99.8)		29.0 (76.0)
614KSH-614KW	52.8 (94.8)		36.4 (64.4)
612YSH-612YW	46.7 (86.3)		8.3 (43.8)
917MSH-917MW	82.6 (98.8)		30.2 (83.4)
0022SH-0022W	30.9 (20.0)		26.6 (67.1)
Average (% of No. of shared reads)	66.1 (84.1)	-	36.1 (72.0)




Phreatobacter
Ochrobactrum
Bosea
Hyphomicrobium
Ancylobacter
Methylocapsa
Roseiarcus
Vampirovibrio
Alsobacter
Pleomorphomonas
Rhizobium
Shinella
Methylocystis
Brucella
Prosthecomicrobium
Bacterial genus
0
2
4
6
LDA score
Water

Genus		*Acinetobacter*	*Brucella*	*Chryseobacterium*	*Clostridium*	*Corynebacterium*	*Haemophilus*	*Moraxella*	*Mycobacterium*	*Nocardia*	*Pseudomonas*	*Staphylococcus*	*Streptococcus*	*Vibrio*
Sample ID	Region	Relative abundance [%]												
425SH	Kinki	–	–	0.03	0.06	0.30	–	1.09	2.13	0.03	–	–	–	–
425W		–	0.46	0.15	–	–	–	–	0.43	–	–	0.03	0.15	–
858SH	Kinki	–	–	–	–	–	–	0.12	–	–	–	–	–	–
858W		–	0.27	–	–	–	–	18.28	0.06	–	–	0.21	0.12	–
513OSH	Kinki	–	–	–	–	3.71	0.12	35.74	–	–	–	0.82	0.82	–
513OW		–	–	–	0.06	–	–	–	0.12	–	–	8.30	0.15	–
94OSH	Kinki	–	–	–	–	–	–	–	–	–	–	0.76	0.49	–
94OW		–	–	–	–	0.03	–	4.01	8.24	–	–	–	–	–
67OSH	Kinki	–	0.58	–	–	–	–	–	–	–	–	–	–	–
67OW		–	1.70	–	–	0.12	0.09	0.85	–	–	–	–	0.03	–
710OSH	Kinki	–	0.03	–	–	1.16	–	0.94	–	–	–	0.15	0.15	–
710OW		–	0.24	–	0.09	3.07	–	5.87	–	–	–	–	0.03	–
712KSH	Kinki	–	0.09	–	–	–	–	–	–	–	–	–	–	–
712KW		–	0.21	–	–	–	–	–	0.06	–	–	–	–	–
820ItSH	Kinki	0.03	–	–	–	5.20	–	22.48	–	–	–	–	–	–
820ItW		–	0.76	–	–	–	–	0.03	–	–	–	–	–	–
852SH	Kanto	–	0.40	–	–	–	–	0.06	–	–	–	0.03	–	–
852W		0.06	0.43	–	–	0.09	–	13.78	0.91	–	0.12	0.43	0.03	–
513SSH	Kanto	–	–	–	–	–	–	0.06	–	–	–	–	–	–
513SW		–	0.12	–	–	–	0.03	0.46	–	–	–	0.30	0.24	–
517ISH	Kanto	–	–	–	–	0.06	–	0.03	–	–	–	–	–	–
517IW		–	0.24	–	0.03	0.06	–	0.33	–	–	–	3.32	–	–
531HMSH	Kanto	–	–	–	–	–	–	0.03	–	–	–	–	–	–
531HMW		–	0.15	–	–	–	–	0.06	0.09	–	–	1.58	–	–
614KSH	Kanto	–	–	–	–	0.24	–	0.79	0.27	–	–	–	–	–
614KW		–	0.27	–	–	–	–	0.85	–	–	0.03	–	0.03	–
612YSH	Kanto	–	–	–	–	–	–	5.66	–	–	–	9.67	–	–
612YW		0.03	0.36	–	0.06	–	–	0.03	0.46	–	–	0.40	–	–
917MSH	Tohoku	0.52	–	0.21	–	10.10	–	1.52	–	–	–	19.95	4.84	0.67
917MW		0.06	0.12	–	0.12	0.79	–	6.75	0.13	–	0.09	32.03	0.03	0.03
0022SH	Tohoku	–	0.03	–	–	0.03	–	2.46	–	–	–	0.58	1.13	0.40
0022W		–	0.33	–	–	–	–	0.06	–	–	–	0.12	0.03	0.12
Mean [%]		0.02	0.21	0.01	0.01	0.78	0.01	3.82	0.40	0.00	0.01	2.46	0.26	0.04
Median [%]		0.00	0.11	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.03	0.00	0.00
Appearance frequency [%]		16	59	9.0	19	44	9.0	81	34	3.0	9.0	53	47	13