



OPEN ACCESS

EDITED BY

Angela Kent,
University of Illinois at Urbana-Champaign,
United States

REVIEWED BY

Scott P. Keely,
United States Environmental Protection
Agency (EPA), United States
Sebastien P. Faucher,
McGill University, Canada

*CORRESPONDENCE

Dominique L. Chaput
✉ dominique.chaput@ggc.scot.nhs.uk
Maria Luisa Ricci
✉ marialuisa.ricci@iss.it

†These authors have contributed equally
to this work

RECEIVED 21 February 2023

ACCEPTED 23 May 2023

PUBLISHED 19 June 2023

CITATION

Scaturro M, Del Chierico F, Motro Y,
Chaldoupi A, Flountzi A, Moran-Gilad J,
Girolamo A, Koutsiomani T,
Krogulska B, Lindsay D, Matuszewska R,
Papageorgiou G, Pancer K, Panoussis N,
Rota MC, Uldum SA, Velonakis E,
Chaput DL and Ricci ML (2023)
Premise plumbing bacterial
communities in four European cities
and their association with *Legionella*.
Front. Microbiomes 2:1170824.
doi: 10.3389/frmbi.2023.1170824

COPYRIGHT

© 2023 Scaturro, Del Chierico, Motro,
Chaldoupi, Flountzi, Moran-Gilad, Girolamo,
Koutsiomani, Krogulska, Lindsay,
Matuszewska, Papageorgiou, Pancer,
Panoussis, Rota, Uldum, Velonakis, Chaput
and Ricci. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Premise plumbing bacterial communities in four European cities and their association with *Legionella*

Maria Scaturro^{1,2}, Federica Del Chierico³, Yair Motro⁴,
Angeliki Chaldoupi⁵, Anastasia Flountzi^{2,5},
Jacob Moran-Gilad^{2,4}, Antonietta Girolamo¹,
Thomai Koutsiomani⁵, Bozena Krogulska⁶, Diane Lindsay^{2,7},
Renata Matuszewska⁶, Georgios Papageorgiou⁵,
Katarzyna Pancer⁶, Nikolaos Panoussis⁵, Maria Cristina Rota¹,
Søren Anker Uldum^{2,8}, Emmanuel Velonakis⁹,
Dominique L. Chaput^{7*†} and Maria Luisa Ricci^{1,2*†}

¹National Reference Laboratory for *Legionella*, Department of Infectious Disease, Istituto Superiore di Sanità, Rome, Italy, ²ESCMID Study Group for *Legionella* Infections (ESGLI), Basel, Switzerland, ³Multimodal Laboratory Medicine Research Area, Bambino Gesù Children's Hospital, IRCCS, Unit of Human Microbiome, Rome, Italy, ⁴Department of Health Systems Management, School of Public Health, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel, ⁵Central Public Health Laboratory, Hellenic National Public Health Organization, Athens, Attiki, Greece, ⁶Department of Virology, National Institute of Public Health, NIH (NIPH-NIH), Warsaw, Poland, ⁷Scottish Microbiology Reference Laboratories, Glasgow, United Kingdom, ⁸Department of Bacteria, Parasites and Fungi, Staten Serum Institute, Copenhagen, Denmark, ⁹Hellenic National School of Public Health, Athens, Greece

Introduction: *Legionella* species are Gram negative, facultative, intracellular bacteria found in natural and engineered water systems. Understanding the bacterial interactions underlying the success of *Legionella* in aquatic environments could be beneficial for control.

Materials and methods: We aimed to profile, by 16S rRNA gene amplicon sequencing (V3-V4), the bacterial communities in premise plumbing systems of buildings in four European cities (Copenhagen, Warsaw, Rome, Athens), and identify positive and negative associations of specific community members to culturable *Legionella*.

Results: The coarse taxonomic composition was similar across the four cities, but Copenhagen and Warsaw had richer, more diverse communities than Athens and Rome, with a greater number of city-specific amplicon sequence variants (ASVs). The cities had statistically significant differences in bacterial communities at the ASV level, with relatively few shared ASVs. Out of 5,128 ASVs, 73 were classified as *Legionella*, and one or more of these were detected in most samples from each city (88.1% overall). Interestingly, the relative abundance of *Legionella* ASVs did not correlate with *Legionella* culture status. Overall, 44.2% of samples were *Legionella* culture positive: 71.4% in Warsaw, 62.2% in Athens, 22.2% in Rome, and 15.2% in Copenhagen. 54 specific ASVs and 42 genera had significant positive or negative associations with culturable *Legionella*. Negative associations included

Staphylococcus, *Pseudomonas*, and *Acinetobacter*. Positive associations included several *Nitrospira* ASVs and one classified as *Nitrosomodaceae* oc32, ASVs in the amoeba-associated genera *Craurococcus-Caldovatus* and *Reyranelia*, and the predatory genus *Bdellovibrio*.

Discussion: Some of these associations are well supported by laboratory studies, but others are the opposite of what was expected. This highlights the difficulties in translating pure culture results to in complex real-life scenarios. However, these positive and negative associations held across the four cities, across multiple buildings and plumbing compartments. This is important because developing better control measures, including probiotic approaches, will require an understanding of ecological relationships that can be generalised across different engineered water systems.

KEYWORDS

microbiome, drinking water system, European countries, *Legionella*, microbial association

1 Introduction

Legionella pneumophila (Lp) can cause Legionnaires' disease, a severe pneumonic illness that, outside the covid-19 pandemic period, has shown a growing trend in cases across European countries (www.ecdc.europa.eu/en/legionnaires-disease/surveillance/atlas). The disease is spread primarily by inhalation of *Legionella*-contaminated aerosols produced by inadequate maintenance of water systems. *Legionella* species are aquatic microorganisms often present at low concentrations in freshwater environments, but through municipal water dissemination, they can reach artificial water systems, where they often proliferate thanks to warmer temperatures and nutrient availability. *Legionella* can live in a planktonic form; but owes its survival to multi-species biofilms and protozoa from which it derives nutrients and protection from physical stress and disinfection procedures (Berry et al., 2006; Pereira et al., 2021; Shaheen and Ashbolt, 2021). *Legionella* abundance is modulated by complex microbial interactions, but even where its relative abundance in a biofilm is low, it can form a significant proportion of the bulk water microbiome through lysis of host amoebal cells and release of *Legionella* (Paniagua et al., 2020).

The survival of Lp in engineered water systems is therefore driven not only by favorable temperature and nutrient conditions but also by interactions with other microbial species found as planktonic cells or in mixed community biofilms. Such is the importance of these other community members, that a probiotic approach, one focused on encouraging competitors, grazers, and antagonists in engineered water systems, has been suggested to control proliferation of Lp (Wang et al., 2013; Cavallaro et al., 2022). Numerous bacterial taxa have been shown in pure culture to have a positive or negative influence on Lp growth. For example, the presence in plumbing systems of *Flavobacterium breve*, a L-

cysteine producing bacterium, was shown to support Lp growth, as Lp cannot produce this amino acid itself (Wadowsky and Yee, 1983). Conversely, several taxa are known to inhibit Lp. *Staphylococcus warneri* produces a peptide called warnericin that has a *Legionella*-specific antagonistic effect (Hécharde et al., 2005). *Pseudomonas* species can inhibit Lp through several mechanisms, including killing host *Acanthamoeba castellanii* (Abd et al., 2008), and production of biosurfactants (Loiseau et al., 2018), toxoflavin (Faucher et al., 2022), bacteriocin-like substances (Guerrieri et al., 2008), and volatile organic compounds with long-range inhibition (Corre et al., 2019). *Bdellovibrio*, an aerobic Gram-negative bacterium known to prey on other Gram-negative taxa, has been shown to cause lysis of at least three *Legionella* species (Tomov et al., 1982). *Bacillus subtilis* can produce a surfactin that inhibits both *Legionella* and its acanthamoebal host (Loiseau et al., 2015). *Neochlamydia* species are obligate intracellular symbionts of amoeba, and their presence blocks infection by *Legionella*, thereby disrupting its replication (Maita et al., 2018). Numerous other taxa have been shown to reduce *Legionella* growth in pure culture, including isolates from the genera *Aeromonas*, *Flavobacterium*, *Acinetobacter*, *Kluyvera*, *Rahnella*, and *Sphingobacterium* (Corre et al., 2019).

Translating culture findings to real-life environmental systems is difficult as two or three isolated species on an agar plate will not reflect the diversity in biofilms or bulk water containing hundreds, if not thousands, of additional taxa. Although they can only show correlation rather than causation. Association studies, where abundance and prevalence of other microbial community members, with different levels of *Legionella* growth, are compared, can indicate which laboratory-based, pure culture findings are most important under natural conditions (i.e. in engineered water systems), pointing to potential targets for probiotic interventions (Cavallaro et al., 2022). A few studies have

used this approach in cooling towers in Quebec, Canada (Paranjape et al., 2020a; Paranjape et al., 2020b) and across the USA (Llewellyn et al., 2017), as well as in premise plumbing of high-rise buildings in New York, USA (Ma et al., 2020).

These studies used microbial community profiling by culture-independent sequencing of bacterial 16S rRNA (Llewellyn et al., 2017; Ma et al., 2020; Paranjape et al., 2020b) or eukaryotic 18S rRNA (Paranjape et al., 2020a) to see which other community members showed positive or negative correlations with *Legionella*. These studies, correlating microbial community taxa with presence of *Legionella*, provided real-world evidence that supports some of the findings from laboratory work with pure cultures. However, the two studies that reported *Legionella*-bacteria associations at the family or genus level were both in cooling towers in North America, so it is unclear whether these associations occur across a wider range of engineered water environments. Therefore, we aimed to determine if similar relationships between *Legionella* and other specific bacterial taxa held across a broad geographical range (four European cities with key differences in water treatment approaches) and across different buildings and premise plumbing compartments, covering a wide temperature range. This required the comparison of the resident bacterial communities in buildings sampled in the four cities. Determining which *Legionella*-community interactions are broadly universal and which are limited to specific systems or conditions will help improve probiotic and other strategies for controlling the growth of *Legionella* in engineered water systems.

2 Results

Of the 204 samples collected across the four cities, 176 were included in the final analyses: 45 from Athens, 45 from Rome, 44 from Copenhagen, and 42 from Warsaw. One of the excluded samples had abnormally low sequencing depth (30 reads remaining after QC), and the rest of the excluded samples fell below the minimum 10 ng DNA threshold required by the sequencing service. Temperature and pH data were collected for a subset of the samples (n=140), and their distributions show some differences among the four cities (Figure S1). However, due to missing measurements and because other potentially important physical/chemical parameters were not measured across all samples (e.g. dissolved carbon and nitrogen content, salts, disinfectant concentration), these parameters were not included as distinct terms in statistical models, but rather they were assumed to be part of what differentiated cities, buildings and compartments, and therefore were implicitly included in these model terms.

In total, 12 390 846 sequences passed QC and were grouped into 5 128 16S rRNA (V3-V4) amplicon sequence variants (ASVs). Sequencing depth per sample ranged from 2 421 to 197 630, with a median depth of 66 159. All but four samples had over 10k sequences, so for analyses where subsampling to equal depth was required, these four samples were excluded in order to increase the amount of sequence data retained across the other samples.

2.1 Main bacterial classes and genera

Across all samples, 44 bacterial phyla were represented in the 16S rRNA V3-V4 amplicon data, of which 10 accounted for at least 1% of sequences (Figure 1A) and 93.0% in total. *Proteobacteria* was the dominant phylum, accounting for 66.1% of all sequences and 2 298 of the 5 128 ASVs. The proteobacterial sequences that could be further resolved all grouped into the classes *Gamma*- and *Alphaproteobacteria*. *Gammaproteobacteria* accounted for 41.4% of sequences and 1 205 ASVs overall and included 39 orders, the dominant one being *Burkholderiales* with 34.4% of total 16S sequences, followed by *Xanthomonadales* and *Pseudomonadales* both at 1.1%. *Gammaproteobacteria* were most abundant in all cities except Athens, where *Alphaproteobacteria* were more abundant (Figure S2). Overall, *Alphaproteobacteria* accounted for 24.7% of sequences and 1 081 ASVs, and included 25 orders, the dominant ones being *Sphingomonadales* (8.6%), *Rhizobiales* 7.4%, *Acetobacterales* (2.3%), *Caulobacteriales* (2.2%), and *Reyranellales* (1.8%).

After *Proteobacteria*, the next most abundant phylum was *Bacteroidota* at 7.8% and 346 ASVs, followed by *Actinobacteriota*, *Acidobacteriota* and *Nitrospirota* at 3.5, 3.3 and 3.1%, respectively (208, 159, and 86 ASVs). While these phyla generally showed similar broad distribution patterns across each city (though with significantly different relative abundances), *Nitrospirota* were noticeably less abundant in Rome, with median relative abundance approximately two orders of magnitude lower than in the other cities (Figure 1A). At the phylum level, 3.3% of sequences remained unclassified, including two of the top 25 most abundant individual ASVs (ASV20 and ASV17, with 0.72% and 0.7% of all sequences, respectively).

The top fifteen most abundant genera are shown in Figure 1B. Their pattern is more variable, with greater spread among samples within cities. For example, the relative abundance of the top genus *Denitratisona*, in the *Gammaproteobacteria* class *Burkholderia*, varied widely, from being barely detectable in some samples, to encompassing over 50% of reads in other samples. Similar differences in relative abundance, several orders of magnitude within the same city, were seen in many of the other top fifteen genera. Although most of these genera were common across the four cities (albeit in significantly different relative abundances), the *Gammaproteobacteria* genera *Blastomonas* and *Craurococcus-Caldovatus* as well as the *Alphaproteobacteria* genus *Porphyrobacter* were nearly absent from Copenhagen samples but relatively abundant elsewhere.

2.2 Bacterial community diversity

Before looking for specific taxa associated with culturable *Legionella*, we examined how the overall bacterial community structure compared across the four cities, and whether it differed depending on *Legionella* culture status of the samples.

Estimated population richness (number of ASVs) was significantly higher and more variable in Copenhagen and

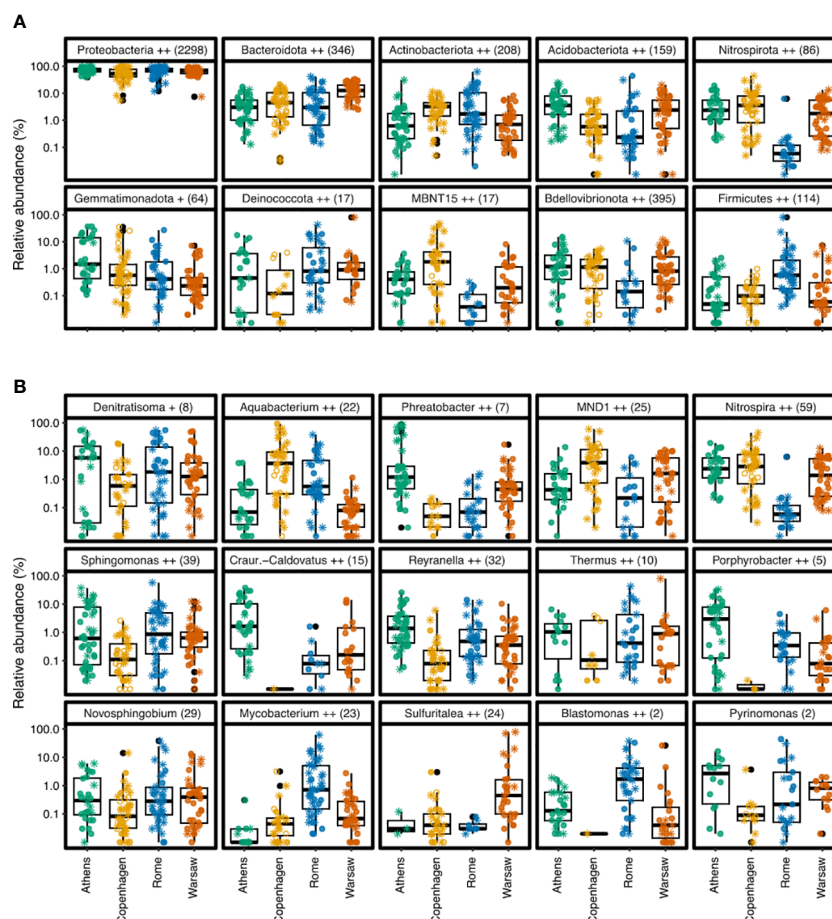


FIGURE 1

Main bacterial phyla and genera in premise plumbing systems in four European cities. (A) Top ten most abundant phyla across all samples. Phyla with significantly different relative abundances among the four cities are (ALDEx2 Kruskal-Wallis test, BH-corrected $p < 0.05$). (B) Top fifteen most abundant genera across all samples. Plus signs next to taxon names indicate phyla and genera with significantly different relative abundances among the four cities (ALDEx2 Kruskal-Wallis test, $p < 0.05$ (+) or $p < 0.01$ (++)). Parentheses next to taxon names show the number of distinct ASVs with that classification. Sequencing depth was normalised to 10 148 sequences per sample, giving a minimum detectable relative abundance of 0.0099%. Filled circles show *Legionella* culture-positive samples, asterisks show culture-negative samples, open circles show samples with no culture data.

Warsaw, compared with Athens and Rome (Figure 2A, with statistical output in Tables S1, S2). Taking into account the influence of the city, culture status also had a significant correlation with richness, with culture-positive samples having, on average, 122.0 more estimated ASVs than culture-negative samples ($p < 0.001$). However, there were significant interactions of city and culture status, indicating that the effect of these factors on richness was not consistent across the four cities. While *Legionella*-positive samples had higher bacterial community richness in Copenhagen, Athens, and, to a lesser extent, in Rome, the opposite was true in Warsaw, where *Legionella*-negative samples had higher richness.

The Shannon diversity index, a measure of entropy, indicates how evenly the reads are distributed among the ASVs in a sample. Communities with a small number of dominant ASVs would have low Shannon indices, whereas more even community structures would result in higher Shannon indices. Here, despite the significant differences in richness among the cities and between culture-positive and culture-negative samples, Shannon index was similar

across samples, indicating a common community structure in these environments (Figure 2B; Tables S3, S4). Neither city ($F_{3,9.99} = 2.7$, $p = 0.1$), culture status ($F_{1,135.41} = 3.7$, $p = 0.057$), or their interaction ($F_{3,127.92} = 1.0$, $p = 0.38$) had a significant influence on Shannon diversity.

PCA on Aitchison distance clearly clustered the samples by their city (Figure 2C). Copenhagen and Warsaw formed looser, broader clusters than Rome and Athens, indicating more intra-city variability (which likely contributes to the much higher overall richness observed in each of those cities). The percent variance explained by the first two components is shown on the axes. PC1 clearly delineates all Copenhagen samples from the others, suggesting a greater difference in the bacterial communities in those samples compared with the other cities. In contrast, PC2 separates all four cities, pointing to the strong influence of local factors at the ASV level.

The most influential ASVs contributing to the separation of Copenhagen samples along PC1 predominantly belong to the *Nitrospira*, with several ASVs within this genus being highly

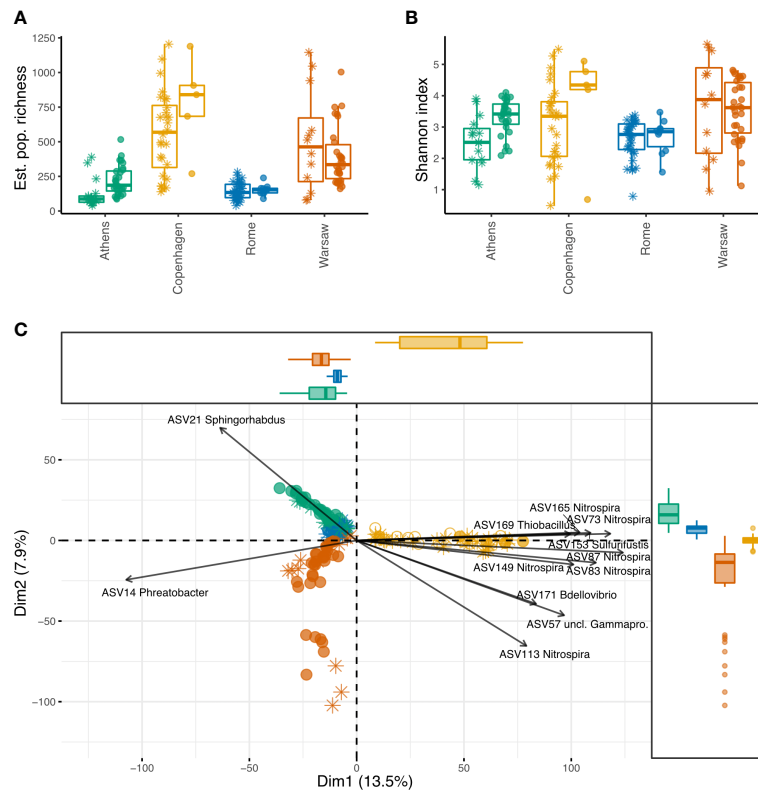


FIGURE 2

Alpha and beta diversity of premise plumbing samples from four European cities, with and without culturable *Legionella*. Asterisks show samples that were *Legionella* culture-negative, filled circles show samples that were *Legionella* culture-positive, open circles show samples without *Legionella* culture data. **(A)** Population richness (number of ASVs) estimated from sample richness using breakaway, which fits nonlinear regression models on the ratios of frequency counts to estimate total richness. This approach does not require sample normalisation, since greater sequencing depth reduces the error of the estimate. City, culture status, and their interaction all had a significant effect, and pairs of cities with significant differences are indicated by plus symbols ($p < 0.001$). **(B)** Sample Shannon diversity index. Samples were normalised to 10 148 sequences. City and culture status were not significant, nor was their interaction. **(C)** Aitchison distance PCA, with arrows indicating the fifteen most influential ASVs. Percent variance explained by the first two components is shown on the axes, and marginal boxplots show the distribution of samples along these components.

correlated (i.e. their arrows follow the same path). Copenhagen stood out as having a higher diversity of *Nitrospira* than the other cities, with 676 ASVs falling into this genus, accounting for 2.6% of its 16S rRNA sequences. In comparison, Warsaw, Athens, and Rome had 257, 97, and 33 *Nitrospira* ASVs, respectively, accounting for 1.1%, 1.4%, and 0.08% of sequences from each of those cities.

We used PERMANOVA, as implemented by the *adonis* function in the *vegan* package, to test the significance of groupings, blocking by building location ID to account for repeated measures (Table S5). The groupings were significantly influenced by both city ($F_{3,157} = 14.3$, $R^2 = 0.2$, $p = 0.0014$) and *Legionella* culture status ($F_{1,157} = 5.7$, $R^2 = 0.027$, $p = 0.0005$), but the interaction of these variables was not statistically significant ($F_{3,157} = 2.7$, $R^2 = 0.037$, $p = 0.054$).

Since PERMANOVA is sensitive to differences in group dispersions, these were tested using the *betadisper* function in the R package *vegan* (Table S6). Group variances were significantly different for city ($F_{3,161} = 15.4$, $p = 0.0001$), indicating that the variability of communities was greater in some cities than in others, which is apparent from the PCA (Figure 2C). The group variances for *Legionella* culture status were not significantly different

($F_{1,163} = 0.077$, $p = 0.78$), so the variability was similar for communities with and without culturable *Legionella*.

Figure 3A shows the intersection of ASVs across the four cities. The rarefied data set used in this analysis (with samples normalised to 10 148 sequences) contained 5086 ASVs. Of these, only 188 (3.7%) were observed in all four cities. Most ASVs, 3545 in total, were detected in only one of the four cities, with 1 827 of these unique to Copenhagen but only 231 ASVs unique to Rome. Not only did the total richness vary markedly between pairs of cities, with Copenhagen and Warsaw having greater richness than Athens and Rome, but the proportion of ASVs unique to each city also differed: 62.9% of ASVs from Copenhagen were not detected anywhere else, whereas these proportions were 47.3% in Warsaw, 34.7% in Athens, and 23.6% in Rome.

2.3 *Legionella* ASVs in the 16S rRNA data

In the rarefied 16S rRNA data set, 73 ASVs were classified as *Legionella* at the genus level, accounting for 0.55% of sequences. Seven of these *Legionella* genus ASVs could be further classified to

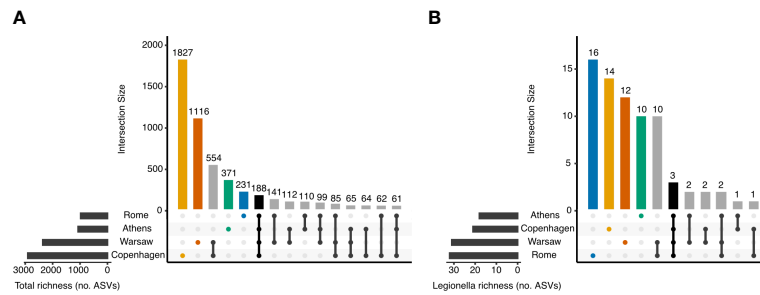


FIGURE 3 Number of shared and distinct ASVs among the four European cities. Samples were normalised to 10 148 sequences prior to determining the ASVs present in each city, and no minimum prevalence or abundance thresholds were used to determine membership. **(A)** All 16S rRNA ASVs, and **(B)** ASVs classified into the genus *Legionella*.

species level. *Legionella* genus ASVs were present in 155 samples overall (88.1%), and in the majority of samples from each city: 77.8%, 95.5%, 91.1% and 88.1% of samples from Athens, Copenhagen, Rome and Warsaw, respectively.

Most of the *Legionella* genus ASVs, 52 out of 73, was unique to one of the four cities. Only three *Legionella* ASVs were detected in all four cities (Figures 3B, 4), and these could all be classified to species level: ASV142 (*Lp*), ASV425 (*L. anisa*), and ASV1223 (*L. waltersii*). Although Rome had the lowest total bacterial ASV richness and the smallest number of unique bacterial ASVs of any city (Figure 3A), it had the largest richness (32 ASVs) and the greatest number of city-specific *Legionella* ASVs (16) (Figure 3B). Overall, 3.3% of ASVs detected in Rome were classified as *Legionella*, whereas this proportion was 1.7, 1.3, and 0.72% in Athens, Warsaw, and Copenhagen, respectively.

The abundance and prevalence of all *Legionella* ASVs present in five or more samples is shown in Figure 4. Most of the *Legionella* ASVs that could be resolved to species level were detected in more than one city, though Copenhagen stood out as rarely having any of these shared, species-resolved ASVs. In many cases, for example in Athens, the specific *Legionella* ASVs detected differed markedly between buildings rather than being distributed evenly across all the samples from that city, and while some buildings (e.g. B1) had a diverse group of *Legionella* ASVs, others (e.g. B2) had only one that was rarely seen elsewhere. A building-specific pattern was less clear in Copenhagen, which had a diverse (albeit low-abundance) population of city-specific ASVs that occurred across all three buildings that were sampled.

Overall, the most abundant *Legionella* genus ASV, ASV142, was further classified as *Lp*, followed by ASV280, which was unclassified

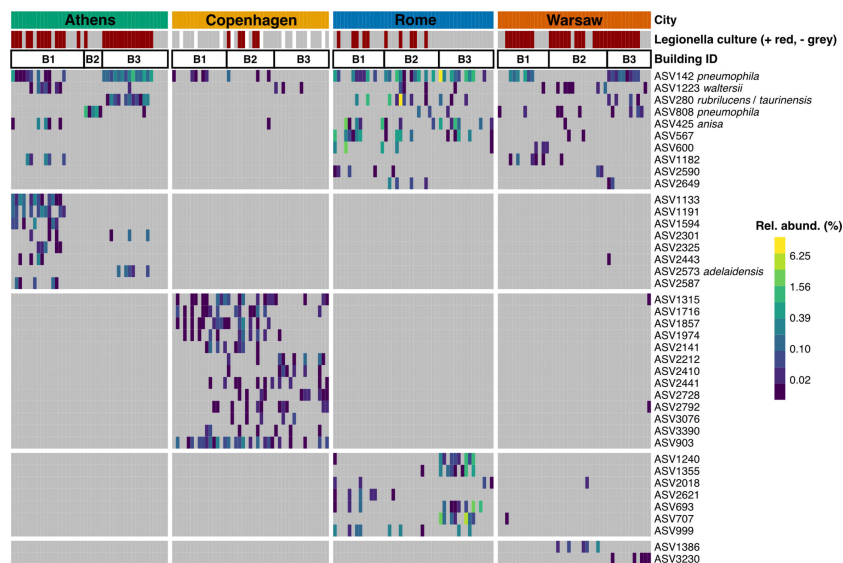


FIGURE 4 Prevalence and relative abundance of ASVs classifying as *Legionella*, across all samples, grouped by city and by building. Samples were normalised to 10 148 sequences. *Legionella* culture status of each sample is shown in the bar at the top (red = positive, grey = negative, white = no culture data collected). ASVs are grouped by their prevalence either across multiple cities (top horizontal block), or predominantly in a single city (bottom four horizontal blocks). *Legionella* genus ASVs that could be further classified have species names attached. Only ASVs present in five or more samples are shown, for clarity. Two rarer ASVs were also further classified to species level: *L. drozanskii* (ASV281), present in four samples (two from Rome, two from Warsaw), and *Lp* (ASV4571), present in one sample from Rome.

at the species level, though BLAST search against GenBank showed it has 100% similarity to both *L. taurinensis* and *L. rubrilucens*. Two rarer ASVs, ASV808 and ASV4571, were also classified as Lp. The latter was detected in a single sample from Rome, but ASV808, which differs from ASV142 by a single nucleotide, was found in all cities except Copenhagen, though it was particularly prevalent in Athens Building 2, which, unlike all other sampled buildings in this study, had no other *Legionella* sequences (Figure 4).

2.4 *Legionella* culture counts versus presence in 16S rRNA data

All samples were tested for culturable *Legionella*, except the first 11 samples from Copenhagen, for which *Legionella* culture counts were not collected. Of the 165 samples with culture data, 73 (44.2%) had culturable *Legionella*. The proportion of *Legionella* culture positive samples differed widely among the four cities: 15.2% from Copenhagen, 22.2% from Rome, 62.2% from Athens, and 71.4% from Warsaw. Culture counts in the *Legionella*-positive samples also varied widely, over five orders of magnitude (Figure 5A). The difference in culturable *Legionella* CFU/L was generally not significant among the four cities, except that Copenhagen had significantly lower CFU/L values than Warsaw in pairwise comparisons (zero-inflated GLMM with Poisson distribution, $t\text{-ratio}_{153} = -3.023$, $p = 0.015$, Tables S7-S9). The building and compartment random effects improved model fit, as did modelling the zero counts separately by City and random effects, suggesting that the distribution of zero values (i.e. *Legionella* culture-negative samples) differed among the cities (as observed in Figure 5A), and also among the buildings and compartments.

In three instances, samples were culture-positive but had no detectable *Legionella* spp. in the sequence data (sequencing depth was normalized to 10 148 sequences per sample, giving a minimum detectable relative abundance of 0.0099%).

Conversely, 75 samples were culture negative but had detectable *Legionella* spp. in the sequence data. Seventeen samples were below the minimum relative abundance and limit of detection of both sequencing and culturing respectively, whereas 70 had detectable *Legionella* in both approaches.

While the detectable relative abundance of *Legionella* spp. sequences in the 16S rRNA data did not differ significantly across all four cities ($F_{3,9.68} = 3.4$, $p = 0.063$), Rome exhibited a different pattern from the others (Figure 5B). In the other cities, the proportion of *Legionella* sequences in each sample did not exceed 1% (with one exception from Athens at 1.06%), but Rome had 16 samples with *Legionella* above this threshold, including 5 samples with *Legionella* sequences exceeded 5%, with one reaching the highest proportion observed in any sample, 17.7%.

Whether the samples were culture-positive or culture-negative had no significant effect on the proportion of sequences that were placed in the genus *Legionella* ($F_{1,107.8} = 0.1$, $p = 0.75$), and this was the case in all cities, as there was no significant interaction between city and presence/absence of culturable *Legionella* ($F_{1,107.8} = 0.1$, $p = 0.75$). Detailed model outputs are shown in Table S10. The random effects terms did not significantly improve the model ($p = 0.51$ and 0.49 for building and compartment ID, respectively), indicating that these factors do not explain a significant amount of the variance in *Legionella* sequence proportions. These findings held regardless of whether the culture data were expressed in binary terms (pos/neg), in discrete categorical terms (log10 bins) or on a continuous scale as CFU/L. For subsequent analyses, only the presence or absence of any culturable *Legionella* was considered, not the CFU/L obtained.

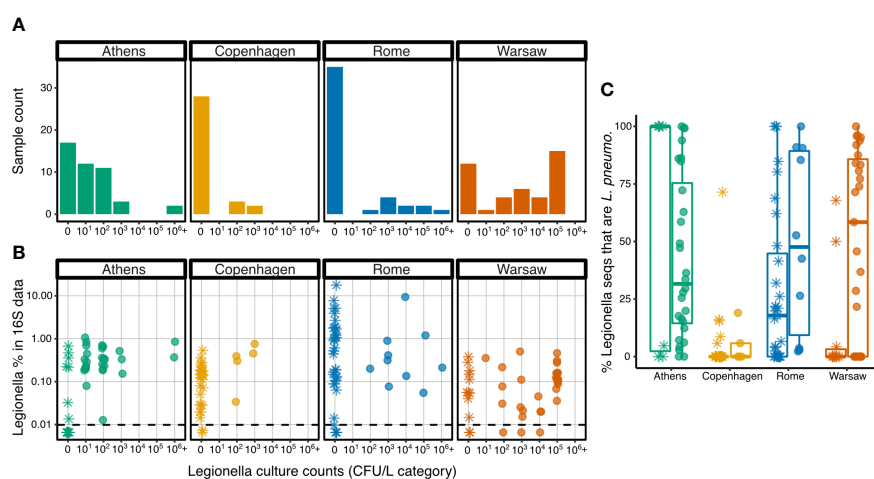


FIGURE 5

Legionella spp. in the culture and 16S rRNA data. (A) Distribution of samples from each city across the culturable *Legionella* categories (log10 bins, with '0' indicating culture-negative samples). (B) Percent 16S rRNA sequences classifying as the genus *Legionella* versus culturable *Legionella* count categories. Dashed horizontal line shows the minimum detectable relative abundance of sequences of the 16S data, normalized to 10 148 sequences per sample, giving a minimum detectable relative abundance of 0.0099%. (C) Percent *Legionella* spp. 16S rRNA sequences that were further classified as Lp. Asterisks indicate samples that were *Legionella* culture-negative, filled circles show samples that were *Legionella* culture-positive.

Since culturing is biased towards Lp, we then examined the proportion of sequences that classified specifically as Lp, rather than all those falling into the genus *Legionella*. As a proportion of all 16S rRNA sequences, there was no significant difference in Lp among the cities or between culture-positive and culture-negative samples (Table S10).

However, differences were observed when the Lp proportion was calculated not from all 16S rRNA sequences, but from those classifying as the genus *Legionella* (Figure 5C). Copenhagen had the lowest proportion of Lp indicating that the majority of *Legionella* sequences detected in that city were from other species (as also seen in Figure 4). In the other cities, some samples had only Lp sequences, some had no Lp sequences among their *Legionella* spp. data, and many had a mix of Lp and other species. Given the variability, the differences among the cities was not significant ($F_{3,8.03} = 1.9, p = 0.21$), but taking into account the influence of the city and the random effects (building and compartment), the culture status of the sample (*Legionella*-positive or negative) had a significant influence on the proportion of *Legionella* genus sequences that were further classified as Lp (Table S11), with the proportion in culture-positive samples being, on average, 24.7% higher than in culture-negative samples.

Despite this agreement between culture status and proportion of *Legionella* sequences classified as Lp, we used the culture status rather than sequence data to delineate between samples with and

without *Legionella* to address our main question: whether other members of the bacterial communities are associated (positively or negatively) with *Legionella*. Sequencing total environmental DNA can detect dead, nonviable, and dormant cells, whereas culturing confirms the presence of viable *Legionella*.

2.5 Associations of bacterial taxa with *Legionella*

Despite the variability in the sample types and the differences in community diversity among the four cities, differential abundance analysis was able to identify 54 specific ASVs (Figure 6A) that are negatively or positively associated with culturable *Legionella*. Only ASVs present across a relatively large number of samples could show significant associations, and given the variability among cities at the ASV level, a relatively small number met this conditions. We therefore also clustered into genera to identify taxa that might fulfill similar roles in different places, despite regional variations in the exact species and strains within those genera. We identified 42 specific genera (Figure 6B) with a significant association, not including those unclassified at the genus level (since the agglomeration step risks grouping together quite different unclassified taxa that share the family level).

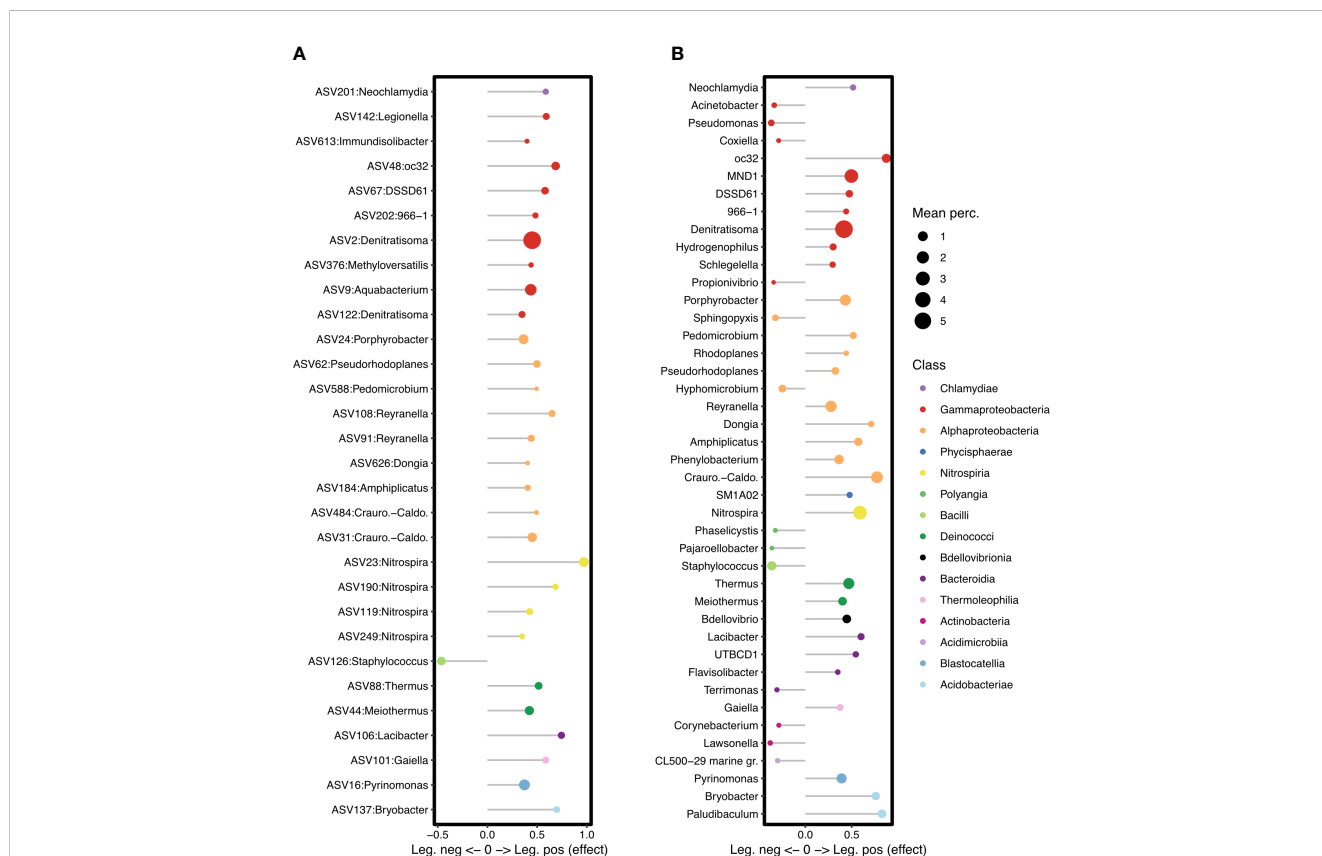


FIGURE 6 Specific ASVs (A) and genera (B) showing a significant association with *Legionella* culture-negative or culture-positive samples. ASVs and genera with effect < 0 are significantly associated with culture-negative samples, and those with effect > 0, with culture-positive samples. Symbols are coloured by Class and scaled to the mean relative abundance of the ASV or genus across all samples.

2.5.1 Negative associations

At the ASV level, only ASV126 was identified as having a significantly negative association with culturable *Legionella*. It was classified as genus *Staphylococcus* (Phylum *Firmicutes*, Class *Bacilli*), but could not be resolved to species level as it showed 100% identity to *S. capitis*, *S. epidermidis*, and *S. capra* when searched against GenBank.

When ASVs were grouped at the genus level, more negative associations were identified, with 13 named genera having a significant negative association with culturable *Legionella* (Figure 6B). These include the *Gammaproteobacteria* genera *Pseudomonas* and *Acinetobacter* (Figure 7A), the *Alphaproteobacteria* genera *Hyphomicrobium* and *Sphingopyxis*, as well as the genus *Staphylococcus*.

2.5.2 Positive associations

Numerous individual ASVs and genera were positively associated with culturable *Legionella* (Figure 6). At the ASV level, ASV142, Lp, had a positive association overall despite its unusual abundance in the culture-negative samples from Rome. Even though the proportion of all *Legionella* genus sequences could not be tied to culturable status, this specific Lp ASV could.

Of the 59 ASVs classified as the nitrite oxidising genus *Nitrospira*, four were positively associated with culturable *Legionella*, and their combined relative abundance across the cities is shown in Figure 7B. Despite the genus *Nitrospira* forming a large proportion of the bacterial communities in Copenhagen, more so than in the other cities, these four specific *Legionella*-associated *Nitrospira* ASVs were hardly detected in samples from there (Figures 7; S3), and they were not among the *Nitrospira* ASVs that contributed to the separation of Copenhagen samples in the

PCA (Figure 2C). The specific *Nitrospira* ASVs that were abundant in Copenhagen were not positively associated with *Legionella*.

At the genus level, 29 genera had a positive association with culturable *Legionella*. These included the genus *Nitrospira* overall, as well as the nitrifying genus *oc32*, from the family *Nitrosomonadaceae*. *Craurococcus* and *Reyranella*, genera known to be associated with amoeba, were also significantly associated with *Legionella* culture-positive samples, as were *Bdellovibrio* and the *Acidobacteria* genera *Paludibaculum* and *Bryobacter*.

3 Discussion

Legionella proliferation in engineered water systems, including premise plumbing, is a wide-scale problem with measurable impacts on human health. Understanding the ecology of *Legionella* in these systems, notably its interactions with other microbial community members, is key to developing more robust and sustainable control measures. Here, we showed that numerous bacterial community members displayed significantly different abundance patterns depending on whether culturable *Legionella* were present or not in a premise plumbing water sample. Taxa that showed a positive association with culturable *Legionella* included members of amoeba-associated genera (*Reyranella* and *Craurococcus-Caldovatus*), those involved in nitrogen cycling (*Nitrospira*, *Denitratisoma*, *Nitrosomonadaceae*), and the predatory genus *Bdellovibrio*. Other taxa, notably *Staphylococcus*, *Pseudomonas*, and *Acinetobacter*, were more abundant in samples without culturable *Legionella*, suggesting a possible inhibitory effect. These positive and negative relationships held across the four European cities and across the different buildings and plumbing compartments. This is important because developing better control

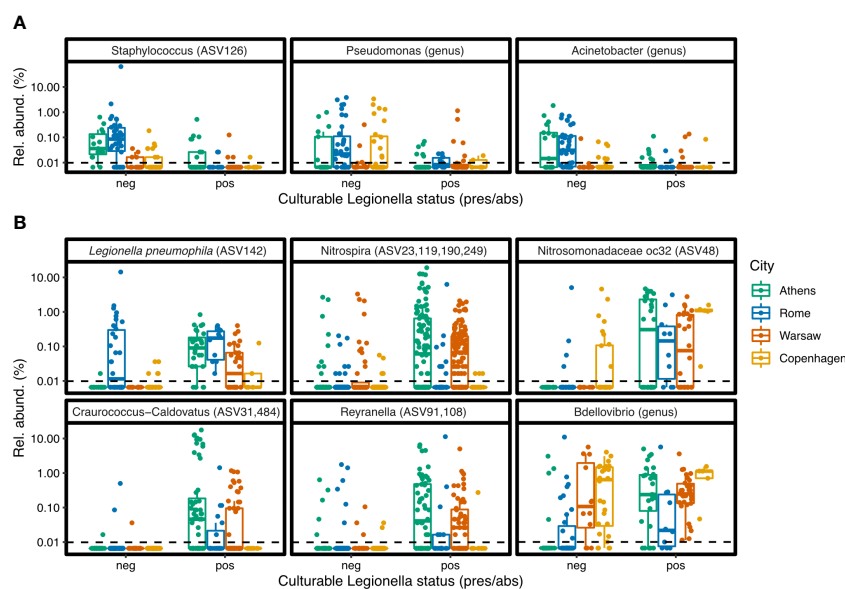


FIGURE 7

Examples of differentially abundant ASVs and genera. (A) Taxa associated with *Legionella* culture-negative samples. (B) Taxa associated with *Legionella* culture-positive samples. Differential abundance analysis was carried out separately on the ASV-level and genus-level data, using ALDEx2.

measures, including probiotic approaches, will require an understanding of ecological relationships that can be generalized across different systems.

The differences between culture-positive and culture-negative samples were not limited to individual bacterial genera and ASVs. At the community structure level, although we found no significant differences in Shannon index between culture positive and negative samples, pointing to similar dominance/evenness structures [also reported in Canadian cooling towers (Paranjape et al., 2020b)], the estimated community richness differed significantly, with higher richness overall in *Legionella* culture-positive samples. The same observation was reported in cooling towers across the USA, where bacterial community richness was positively correlated with detectable *Legionella* (Llewellyn et al., 2017). The USA premise plumbing study limited its analysis to the Class level, so specific families or genera associated with *Legionella* could not be resolved, but both cooling tower studies showed that *Pseudomonas* (or the family level, *Pseudomonadaceae*, in the USA study) had a negative correlation to *Legionella* spp., suggesting that the inhibitory effect seen in laboratory studies might also be occurring in cooling towers. The only other taxon that showed a negative relationship specifically to Lp, reported in the Quebec study, was the genus *Sphingobium*. The same study reported that numerous genera were enriched in *Legionella*-positive samples, including *Porphyrobacter*, *Sediminibacterium*, *Sphingopyxis*, *Hyphomicrobium*, and *Reyranelia*, among others (Paranjape et al., 2020b).

Why greater community richness is associated with a higher probability of growing *Legionella* is unclear. It could point to a lower concentration of residual disinfectant in those compartments, allowing not only *Legionella* but other taxa to flourish. This would not, however, explain the difference observed in Copenhagen, since no disinfection is used there but *Legionella*-positive samples had, on the whole, richer communities than *Legionella*-negative ones. Perhaps the greater community richness in culture-positive samples is due to *Legionella*'s requirement for specific amino acids, which might be in greater supply in richer communities, where several different taxa can supply them. However, the effect of *Legionella* culture status on richness in our study was complicated by a significant interaction with the city, meaning the effect was not observed equally across all cities. Indeed, it was reversed in Warsaw, where culture-negative samples had higher estimated richness.

The differential abundance signal from positively and negatively associated taxa was detectable despite there being significant differences in premise plumbing bacterial communities across the four cities, with relatively little overlap in ASVs, most of which were city-specific. Total richness and number of city-specific ASVs were highest in Copenhagen and Warsaw, but much lower in Athens and Rome. Drinking water systems without a disinfectant residual are known to be richer than those with a residual (Bautista-de los Santos et al., 2016; Thom et al., 2022), so the absence of chemical disinfection in Copenhagen's public water supply likely explains why it had the richest premise plumbing communities, the largest number of city-specific ASVs, the greatest dispersion among samples, and the greatest beta diversity separation from the other cities, accounting for 13.5% of variance across the data set. A possible reason of such high richness of Warsaw compared with

Athens and Rome, although the water is treated with chlorine dioxide, is that it is supplied originally from the Vistula river and the Debe Dam lake that may themselves show a rich microbial diversity (Luo et al., 2020; Deng et al., 2021).

The coarse taxonomic composition of premise plumbing communities was similar in these four cities, with strong dominance of *Proteobacteria* classes *Gamma*- and *Alphaproteobacteria*. This is in line with other studies of water distribution systems (Bruno et al., 2018; Paduano et al., 2020; Goraj et al., 2021; Thom et al., 2022). At finer taxonomic resolution, there were some noticeable differences among the four cities. *Craurococcus-Caldovatus* and *Blastomonas* were almost absent from Copenhagen but relatively abundant in the other cities. Conversely, *Nitrospira* abundance and diversity were much higher in Copenhagen than elsewhere, and numerous *Nitrospira* taxa were responsible for the separation of the Copenhagen samples from the others in beta-diversity analyses. A study conducted in Poland showed overgrowth of *Nitrospira* species in cast iron pipes (Goraj et al., 2021), but this material was not present in the water systems of the Copenhagen buildings that we sampled, which used stainless steel, galvanized iron or copper. This suggests that factors other than pipe material can also lead to differential abundances of microbial taxa such as *Nitrospira*.

Furthermore, the abundance and diversity of sequences classifying as *Legionella* varied across the four cities. Rome had the highest number of *Legionella* ASVs overall, the highest number of city-specific *Legionella* ASVs, and the highest proportions of *Legionella* in the 16S rRNA sequence data, with many samples exceeding 1% (with one sample having 17% of sequences classified as *Legionella*). Additionally, many of the relatively abundant *Legionella* ASVs in Rome were named taxa, including Lp, which are more likely to be involved in human disease than rare, uncharacterised species. In contrast, most *Legionella* ASVs detected in Copenhagen were at low relative abundance, were unnamed and uncharacterised, and were city-specific. There was a near absence of the main named species that cause most human illness. The *Legionella* ASV findings in these two cities sit in contrast to the total community richness and diversity, which was low in Rome and high in Copenhagen. This suggests that rich bacterial communities provide some competition against or dilution of *Legionella* species, but this would contradict our finding that samples with richer total communities were more likely to have culturable *Legionella* (albeit not in all cities).

This contradiction could be due to the absence of a correlation between culturable *Legionella* and abundance in the 16S rRNA sequence data. A similar mismatch between *Legionella* as detected by PCR versus in pure culture was also observed in US cooling towers (Llewellyn et al., 2017). In our study, this mismatch was especially evident in Rome, which had a relatively high abundance of *Legionella* 16S rRNA sequences, even in samples that failed to grow *Legionella* (which included the sample with 17% *Legionella* sequences). Why such a mismatch was observed is unclear; it could be due to disinfection regimes that kill cells after a period of proliferation, without destroying their DNA, or that cause a switch to a viable but nonculturable state, meaning they would be detected by molecular but not culture methods. It could also be due

to the LOD of culturing, 100 CFU/L, or to the fact that culturing can miss *Legionella* that are inside amoeba (whereas these would be detected by DNA-based methods). However, culturing is the gold standard, and detection of *Legionella* DNA is not necessarily a good indicator, since it cannot distinguish between loss of viability and viable but nonculturable state (Llewellyn et al., 2017). For our differential abundance analyses, culturable *Legionella* status, not presence in sequence data, was therefore used to identify negative and positive associations of other community members.

Several of the negative associations that we identified are well supported by laboratory studies. For example, we found that a specific *Staphylococcus* sequence (ASV126) as well as the genus *Staphylococcus* overall were enriched in *Legionella* culture-negative samples, and pure culture experiments have shown that members of this genus can produce anti-*Legionella* peptides (Hécharde et al., 2005; Marchand et al., 2011). However, this negative association was not reported in the cooling tower studies, as *Staphylococcus* species were not detected there (Llewellyn et al., 2017; Paranjape et al., 2020b). The negative effect of *Pseudomonas* on *Legionella* proliferation is well documented from laboratory studies, with *Pseudomonas* species using numerous inhibitory mechanisms (Abd et al., 2008; Guerrieri et al., 2008; Loiseau et al., 2018; Corre et al., 2019; Faucher et al., 2022). The cooling tower association studies in Canada (Paranjape et al., 2020b) and across the USA (Llewellyn et al., 2017) also found that the genus *Pseudomonas* and the family *Pseudomonadaceae*, respectively, had a negative correlation with *Legionella*, and we detected the same pattern on another continent and in premise plumbing compartments, which are different types of engineered water systems. This suggests that the inhibitory effects of *Pseudomonas* species on *Legionella* transcend the laboratory and can be generalised across multiple types of engineered water systems. The negative relationship in cooling towers was thought to be due (at least in part) to chlorination, which kills *Legionella* but allows *Pseudomonas* to proliferate (Paranjape et al., 2020b). However, this explanation does not entirely fit our data, because in Copenhagen, there is no chlorination but the observation still holds (there were more *Pseudomonas* in culture-negative samples).

While the *Staphylococcus* and *Pseudomonas* negative associations were supported by published laboratory and/or cooling tower association studies, other negative relationships identified in our data were the opposite of what was reported in cooling towers. *Sphingopyxis* and *Hyphomicrobium* (both *Alphaproteobacteria*) were positively associated with *Legionella* in cooling towers (Paranjape et al., 2020b) but negatively associated in our data. These genera are particularly abundant in biofilms and are thought to help in biofilm formation (Vaz-Moreira et al., 2011; Ren et al., 2015), so an interesting negative association which may relate to the more organically rich environments of cooling towers and may be less pertinent in nutrient poorer premise plumbing systems that are continuously flushed. In addition, based on laboratory experiments, *Bdellovibrio* should have had a negative influence on *Legionella* proliferation, but was instead positively associated with culturable *Legionella*. *Bdellovibrio* is a predatory taxon but is not a specific grazer of *Legionella* (Cavallaro et al., 2022), and it has shown selective predation when cultured with two predation

sensitive bacteria simultaneously (Rogosky et al., 2006), so perhaps it was instead grazing on *Legionella*'s competitors in the biofilm. Another possible explanation is proteolytic degradation of prey bacteria in multispecies cultures that results in the release of nutrients, allowing some prey organisms, expected to be reduced, to instead proliferate in the presence of *Bdellovibrio bacteriovorus* (Hobley et al., 2006). This highlights that real-world effects are more complex than simple laboratory experiments. Seeding systems with *Bdellovibrio* to control *Legionella*, for example, could have unintended consequences if complex ecological interactions are poorly understood.

While the positive association of *Bdellovibrio* was the opposite of what some laboratory experiments would have suggested, other positive associations were easier to explain. Several were taxa known to be associated with amoebae, notably species in the *Craurococcus-Caldovatus* genus grouping (Thomas et al., 2006) and in *Reyranella*. The positive association of these genera with culturable *Legionella* likely points to a higher abundance of amoebae in general, and thus a greater availability of host cells for *Legionella* growth and replication. A positive association of *Reyranella* was also reported in the Canadian cooling tower study (Paranjape et al., 2020b). Another amoeba-associated taxon, the symbiont *Neochlamydia*, also showed a positive association with *Legionella*; however, this correlation is more puzzling, as *Neochlamydia* species can confer resistance of their host amoebal cell to infection by *Legionella* and would be expected to have a negative influence (Cavallaro et al., 2022). In our data, this positive association might simply be highlighting an abundance of amoeba in general.

Several other taxa that were positively associated with culturable *Legionella* appear to be important in nitrogen cycling, which was not apparent in the cooling tower association studies (Llewellyn et al., 2017; Paranjape et al., 2020b). A number of ASVs in the nitrite-oxidising genus *Nitrospira*, as well as the genus overall, showed a strong positive association. This was despite the abundance and diversity of *Nitrospira* ASVs in Copenhagen samples, which only rarely grew *Legionella*. The specific *Nitrospira* ASVs that were most influential in separating the Copenhagen samples from the other cities in the PCA were not the same *Nitrospira* ASVs that had a positive association with *Legionella*. This suggests that even within the same genus, taxa (species, subspecies) can have different types of community interactions, with some species exhibiting a positive association with *Legionella* and others not. In the case of *Nitrospira* at least, but perhaps more generally, the genus level may be too coarse for differential abundance analysis and could mask contrasting patterns at finer taxonomic levels.

Of course, association studies like ours show only correlation, not causation. The taxa with significant differential abundances are not necessarily supporting the proliferation or inhibiting the growth of *Legionella* directly. These taxa, including *Legionella*, could all be responding to other local conditions, for example temperature or disinfectant regime. The scale must also be considered in studies like this; relationships that are statistically significant across cities do not necessarily hold in individual biofilms. Finally, while culture data are quantitative (counts per L), 16S rRNA sequence data can only be semi-quantitative, as they are compositional in nature. They

indicate the proportion of the bacterial community accounted for by each ASV, but do not give any indication of the total microbial biomass in these samples, which would have had to have been measured in parallel using a different method (e.g. flow cytometry, qPCR), or estimated using a spike-in control, neither of which we included here but which would be useful to consider in future studies of this nature.

In conclusion we found that significant associations (negative and positive) of specific bacterial taxa with culturable *Legionella* could be identified across a highly variable set of samples covering different premise plumbing compartments, different buildings, in four cities across Europe. Some of these associations were previously reported in cooling towers in North America and are supported by laboratory experiments with pure cultures. Our study suggests some of these associations are therefore consistent across many different environments where *Legionella* is found. Identifying ecological associations that transcend the laboratory and apply to different engineered water systems is important for developing new approaches to *Legionella* control.

4 Materials and methods

4.1 Sampling locations and sample collection

Four European capital cities were included in this study: Athens (Greece), Copenhagen (Denmark), Rome (Italy), and Warsaw (Poland). In 2015, Italy and Denmark had population LD incidence rates of >2 cases per 100 000, whereas the incidence rates in Greece and Poland were 0.01-0.49 cases per 100 000 (Jong and Hallström, 2022). The four cities differ in their water treatment approaches, the most notable difference being that Copenhagen, unlike the other three cities, does not use chlorination in its drinking water distribution system.

From each city, three buildings were selected among those with larger water systems; this included a mix of hospitals, hotels, healthcare facilities, and public buildings, some of which had additional water treatment systems in place. Within each building, four premise plumbing compartments were chosen among those mentioned by international guidelines as more favorable to *Legionella* proliferation (with the exception of municipal water taken as control), for repeated sampling: municipal water at the entrance of each building (MW), boiler bottom (BB) or after heater exchanger (AHE), hot recirculating water (HRW), and the farthest point from the boiler or heat exchanger (FPB). Repeated samples of each compartment were collected four (Athens, Copenhagen, Rome) or five (Warsaw) times between winter 2016-2017 and spring 2018.

Two litres of water were collected without flushing, in order to represent the worst condition of water system's contamination, from each compartment at each sampling point. One litre was concentrated to 10 mL using sterile 0.22 µm polycarbonate filter membrane for *Legionella* enumeration by culture, following the ISO 11731:2017 protocol, for which a limit of detection of 100 CFU/L was established. The second litre was filtered through a sterile 0.22

µm polycarbonate membrane, which was used for DNA extraction. Water temperature and pH were measured in a subset of samples.

4.2 DNA extraction and 16S rRNA amplicon sequencing

Genomic DNA was extracted using the PowerWater DNA isolation kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Briefly, the filter membrane was cut into pieces with ethanol and flame-sterilised scissors, then inserted into a bead beating tube containing lysis buffer. After vortexing the cells for 2 minutes, lysates were transferred to a DNA-retaining spin column for the purification steps, performed in a QIAcube instrument. DNA was eluted in 100 µl elution buffer (10 mM Tris pH 8.5).

High-throughput sequencing of the bacterial 16S rRNA gene V3-V4 hypervariable region was performed by a commercial genomics service (Eurofins Genomics, <https://www.eurofinsgenomics.eu>). V3-V4 amplicons were generated using the Eurofins standard primers V3V4-F (TACGGGAGGCAGCAG) and V3V4-R (CCAGGGTATCTAATCC). Libraries were cleaned, quantified and pooled for sequencing on the Illumina MiSeq platform with v3 chemistry (2x300 bp).

4.3 Bioinformatics processing

Demultiplexed fastq files obtained from Eurofins were first grouped by the instrument and run number specified in the sequence identifier lines, to account for the different sequencing error profiles that can occur on separate instruments and runs. DADA2 v1.16.0 (Callahan et al., 2016) was used for quality control, error correction, taxonomic assignment, and generation of an amplicon sequence variant (ASV) table. Briefly, sequences were truncated based on their quality profiles, with settings specific to each sequencing run, and filtered to remove those with ambiguous and low-quality bases (full details in the R scripts available on GitHub). After the error modelling step, sample inference was run with the pool=pseudo option to maximise detection of rare variants while remaining within the limits of our computing resources. Paired reads were then merged, discarding any that did not overlap by at least 25 bases with no mismatches. With this primer set, merged reads show a bimodal length distribution with modes at 403 and 427/428 base pairs. Sequences outside the size range 400-430 bp were removed. ASV tables from the different sequencing runs were merged, chimeras were identified and removed using the "consensus" method in DADA2, and taxonomy was assigned using DADA2's implementation of the native Bayesian classifier (Wang et al., 2007), against the Silva NR database, v1.38 (Quast et al., 2012; Glöckner et al., 2017). Further classification to species level was attempted using the Silva v1.38 species training set.

The R package phyloseq (McMurdie and Holmes, 2013) was used for further pruning and filtering to remove likely artefacts and non-target sequences. Sequences classifying as Archaea, Eukarya, or unclassified at the Domain level were removed, as were those identified as mitochondria or chloroplasts. A prevalence filter was

applied to keep only ASVs observed in at least three samples, and samples with fewer than 2000 reads were discarded.

4.4 Statistical analysis

For most analyses, we used compositional methods that do not require normalisation of sequencing depth (Gloor et al., 2016). However, where normalisation was required, e.g. sample Shannon index, comparison of ASV intersections, and proportions of taxonomic groups, we subsampled (with replacement) down to 10 148 sequences per sample, discarding the samples smaller than this in order to maximise the amount of data retained.

Linear and generalised linear mixed effects models [R packages lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), and glmmTMB (Brooks et al., 2017)] were used to account for the hierarchical and repeated measures structure of the data, and for the zero-inflated Poisson distribution observed with the *Legionella* culture count data. City, and where relevant, presence/absence of culturable *Legionella*, were specified as fixed effects. The repeated measurements within each compartment were accounted for by modelling building compartment as a random effect, nested within the building ID, since compartments within a building might have more similar microbiomes due to a common water source and local disinfection regime. The function ranova in the lmerTest package was used to compute ANOVA-like tables to assess whether these random effect terms improved model fit, and pairwise comparisons among the four cities were computed from the model output with the lsmeans package (Lenth, 2016).

For alpha diversity analyses, population richness (the number of ASVs in the populations from which samples were taken, rather than the number of ASVs in the samples themselves) was estimated using the R package breakaway (Willis and Bunge, 2015), and hypothesis testing was carried out with the functions beta and beta_random (Willis et al., 2017), which account for the uncertainties in the richness estimates. Fixed and random effects were as above. Sample Shannon index (at the ASV level) was calculated in phyloseq. The intersection of ASVs among the four cities was plotted using the R package UpSetR (Gehlenborg, 2019).

For beta diversity analyses, we used the compositional Aitchison distance (Gloor et al., 2016), equal to the Euclidean distance calculated from a center-log-ratio (CLR) transformed ASV table. Prior to the CLR transform, which cannot be done on zero values, we applied the count zero multiplicative approach implemented in the zCompositions package (Palarea-Albaladejo and Martín-Fernández, 2015). Samples and taxa were ordinated by PCA and plotted using the factoextra package (Kassambara and Mundt, 2020). The influence of city and presence/absence of culturable *Legionella* (CFU_pres_abs) was assessed by permutational multivariate analysis of variance (PERMANOVA), as implemented by the adonis function of the vegan package (Oksanen et al., 2020). Compartment ID was specified as a blocking variable (strata) to account for repeated measures. Multivariate homogeneity of group dispersions was tested for city and *Legionella* pres/abs with the betadisper function.

Differential abundance analysis was carried out on the ASV, genus-agglomerated, and phylum-agglomerated tables using the

compositional approach implemented in ALDEx2 (Fernandes et al., 2013; Fernandes et al., 2014). To assess the differential abundance of phyla and genera among the four cities, the ANOVA module in ALDEx2 was used to perform a multi-group Kruskal-Wallis test, with Benjamini-Hochberg corrected p-values. To identify ASVs and genera that were associated with culturable *Legionella* across all samples, presence/absence of culturable *Legionella* was specified as the fixed effect, and significance testing used a Wilcoxon Rank Sum test and Welch's t-test, with Benjamini-Hochberg corrected p-values.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/ena>, https://github.com/dlchpt/Legionella_microbiome_paper_scripts.

Author contributions

MLR and DC conceptualized, supervised the study and wrote the paper. DC provided also the conclusive data analyses. MS designed sample collection scheme and provided laboratories procedures supervision. YM and JM-G provided the first raw data analyzed by FD. DL provided comments, review and editing. AC, AF, AG, TK, BK, RM, GP, KP, NP, SU, EV performed sample collection and analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by ESCMID as the first grant provided to ESGLI (ESCMID Study group of *Legionella* Infections) in 2016.

Acknowledgments

We are grateful to the two hospital managers of San Camillo Hospital, Rome, Dr. Daniela Orazi and Dr. Cesira Natalina Piscioneri for their kind availability in supporting us in this study, as well as Anna Rita Sanzarello, Stefano Polidori for their precious support during the samplings at the hospital's water system. We thank Dr. Alessio Lanni for his kind support in sampling and analysis of the water samples. We wish to thank Dr. Rodrigo Bacigalupe from University of Edinburgh, for his critical evaluation of the paper. We are also grateful to Mrs Despina Karamousa from the Region of Attica, Regional Unit of the Central Athens Sector, Department of Health and Environmental Control and Mrs Aspasia Gatsiou from the Specialized Oncologic Hospital of Piraeus "METAXA," for their valuable collaboration, continuous support and commitment to the project and the laboratory. Finally, we are grateful to the many individuals who have taken the time to publish open-access tutorials, bioinformatics workflows, scripts, R packages, and other resources for the benefit of the scientific community.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frmbi.2023.1170824/full#supplementary-material>

References

- Abd, H., Wretling, B., Saeed, A., Idsund, E., Hulthenby, K., and Sandström, G. (2008). *Pseudomonas aeruginosa* utilises its type III secretion system to kill the free-living amoeba *Acanthamoeba castellanii*. *J. Eukaryot. Microbiol.* 55, 235–243. doi: 10.1111/j.1550-7408.2008.00311.x
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67 (1), 1–48. doi: 10.18637/jss.v067.i01
- Bautista-de los Santos, Q. M., Schroeder, J. L., Sevillano-Rivera, M. C., Sungthong, R., Ijaz, U. Z., Sloan, W. T., et al. (2016). Emerging investigators series: microbial communities in full-scale drinking water distribution systems – a meta-analysis. *Environ. Sci.: Water Res. Technol.* 2, 631–644. doi: 10.1039/C6EW00030D
- Berry, D., Xi, C., and Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Curr. Opin. Biotechnol.* 17, 297–302. doi: 10.1016/j.copbio.2006.05.007
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* 9, 378. doi: 10.32614/RJ-2017-066
- Bruno, A., Sandionigi, A., Bernasconi, M., Panio, A., Labra, M., and Casiraghi, M. (2018). Changes in the drinking water microbiome: effects of water treatments along the flow of two drinking water treatment plants in a urbanized area, Milan (Italy). *Front. Microbiol.* 9. doi: 10.3389/frmbi.2018.02557
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Cavallaro, A., Rhoads, W. J., Huwiler, S. G., Stachler, E., and Hammes, F. (2022). Potential probiotic approaches to control *Legionella* in engineered aquatic ecosystems. *FEMS Microbiol. Ecol.* 98 (8), fiac071. doi: 10.1093/femsec/fiac071
- Corre, M.-H., Delafont, V., Legrand, A., Berjeaud, J.-M., and Verdon, J. (2019). Exploiting the richness of environmental waterborne bacterial species to find natural *Legionella pneumophila* competitors. *Front. Microbiol.* 9. doi: 10.3389/frmbi.2018.03360
- Deng, R., Chen, X., Qiu, L.-P., Chen, J.-Z., and Meng, S.-L. (2021). Bacterial community structure and diversity in the aqueous environment of shihou lake and its relationship with environmental factors. *Indian J. Microbiol.* 61, 475–486. doi: 10.1007/s12088-021-00974-y
- Faucher, S. P., Matthews, S., Nickzad, A., Vounba, P., Shetty, D., Bédard, É., et al. (2022). Toxoflavin secreted by *Pseudomonas alcaliphila* inhibits growth of *Legionella pneumophila* and its host *Vermamoeba vermiformis*. *Microbiology* 216, 118328. doi: 10.1101/2022.01.08.475489
- Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., and Gloor, G. B. (2013). ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-seq. *PLoS One* 8, e67019. doi: 10.1371/journal.pone.0067019
- Fernandes, A. D., Reid, J. N., Macklaim, J. M., McMurrrough, T. A., Edgell, D. R., and Gloor, G. B. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2, 15. doi: 10.1186/2049-2618-2-15
- Gehlenborg, N. (2019) *UpSetR: a more scalable alternative to venn and euler diagrams for visualizing intersecting sets*. Available at: <https://CRAN.R-project.org/package=UpSetR>.
- Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., et al. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi: 10.1016/j.jbiotec.2017.06.1198
- Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., and Egozcue, J. J. (2016). It's all relative: analyzing microbiome data as compositions. *Ann. Epidemiol.* 26, 322–329. doi: 10.1016/j.annepidem.2016.03.003
- Gojaj, W., Pytlak, A., Kowalska, B., Kowalski, D., Grządziel, J., Szafranek-Nakonieczna, A., et al. (2021). Influence of pipe material on biofilm microbial communities found in drinking water supply system. *Environ. Res.* 196, 110433. doi: 10.1016/j.envres.2020.110433
- Guerrieri, E., Bondi, M., Sabia, C., de Niederhäusern, S., Borella, P., and Messi, P. (2008). Effect of bacterial interference on biofilm development by *Legionella pneumophila*. *Curr. Microbiol.* 57, 532–536. doi: 10.1007/s00284-008-9237-2
- Héchar, Y., Ferraz, S., Bruneteau, E., Steinert, M., and Berjeaud, J.-M. (2005). Isolation and characterization of a *Staphylococcus warneri* strain producing an anti-legionella peptide. *FEMS Microbiol. Lett.* 252, 19–23. doi: 10.1016/j.femsle.2005.03.046
- Hobley, L., King, J. R., and Sockett, R. E. (2006). *Bdellovibrio* predation in the presence of decoys: three-way bacterial interactions revealed by mathematical and experimental analyses. *Appl. Environ. Microbiol.* 72, 6757–6765. doi: 10.1128/AEM.00844-06
- Jong, B., and Hallström, L. P. (2022). European Surveillance of legionnaires' disease. *Curr. Issues Mol. Biol.* 42, 81–96. doi: 10.21775/cimb.042.081
- Kassambara, A., and Mundt, F. (2020) *Factoextra: extract and visualize the results of multivariate data analyses*. Available at: <https://CRAN.R-project.org/package=factoextra>.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Soft.* 82 (13), 1–26. doi: 10.18637/jss.v082.i13
- Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *J. Stat. Soft.* 69 (1), 1–33. doi: 10.18637/jss.v069.i01
- Lewellyn, A. C., Lucas, C. E., Roberts, S. E., Brown, E. W., Nayak, B. S., Raphael, B. H., et al. (2017). Distribution of *Legionella* and bacterial community composition among regionally diverse US cooling towers. *PLoS One* 12, e0189937. doi: 10.1371/journal.pone.0189937
- Loiseau, C., Portier, E., Corre, M.-H., Schlüsselhuber, M., Depayras, S., Berjeaud, J.-M., et al. (2018). Highlighting the potency of biosurfactants produced by *Pseudomonas* strains as anti-*Legionella* agents. *BioMed. Res. Int.* 2018, 1–15. doi: 10.1155/2018/8194368
- Loiseau, C., Schlüsselhuber, M., Bigot, R., Bertaux, J., Berjeaud, J.-M., and Verdon, J. (2015). Surfactin from *Bacillus subtilis* displays an unexpected anti-*Legionella* activity. *Appl. Microbiol. Biotechnol.* 99, 5083–5093. doi: 10.1007/s00253-014-6317-z
- Luo, X., Xiang, X., Yang, Y., Huang, G., Fu, K., Che, R., et al. (2020). Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. *FEMS Microbiol. Ecol.* 96, fiae132. doi: 10.1093/femsec/fiae132
- Ma, X., Pierre, D., Bibby, K., and Stout, J. E. (2020). Bacterial community structure correlates with *Legionella pneumophila* colonization of new York city high rise building premises plumbing systems. *Environ. Sci.: Water Res. Technol.* 6, 1324–1335. doi: 10.1039/C9EW01084J
- Maita, C., Matsushita, M., Miyoshi, M., Okubo, T., Nakamura, S., Matsuo, J., et al. (2018). Amoebal endosymbiont *Neochlamydia* protects host amoebae against *Legionella pneumophila* infection by preventing *Legionella* entry. *Microbes Infect.* 20, 236–244. doi: 10.1016/j.micinf.2017.12.012
- Marchand, A., Verdon, J., Lacombe, C., Crapart, S., Héchar, Y., and Berjeaud, J. M. (2011). Anti-*Legionella* activity of staphylococcal hemolytic peptides. *Peptides* 32, 845–851. doi: 10.1016/j.peptides.2011.01.025
- McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8, e61217. doi: 10.1371/journal.pone.0061217
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al. (2020) *Vegan: community ecology package*. Available at: <https://CRAN.R-project.org/package=vegan>.

- Paduano, S., Marchesi, I., Casali, M. E., Valeriani, F., Frezza, G., Vecchi, E., et al. (2020). Characterisation of microbial community associated with different disinfection treatments in hospital hot water networks. *IJERPH* 17, 2158. doi: 10.3390/ijerph17062158
- Palarea-Albaladejo, J., and Martín-Fernández, J. A. (2015). zCompositions [[/amp]] mdash; r package for multivariate imputation of left-censored data under a compositional approach. *Chemometr Intell. Lab. Syst.* 143, 85–96. doi: 10.1016/j.chemolab.2015.02.019
- Paniagua, A. T., Paranjape, K., Hu, M., Bédard, E., and Faucher, S. P. (2020). Impact of temperature on *Legionella pneumophila*, its protozoan host cells, and the microbial diversity of the biofilm community of a pilot cooling tower. *Sci. Total Environ.* 712, 136131. doi: 10.1016/j.scitotenv.2019.136131
- Paranjape, K., Bédard, É., Shetty, D., Hu, M., Choon, F. C. P., Prévost, M., et al. (2020a). Unravelling the importance of the eukaryotic and bacterial communities and their relationship with legionella spp. ecology in cooling towers: a complex network. *Microbiome* 8 (1), 157. doi: 10.1186/s40168-020-00926-6
- Paranjape, K., Bédard, É., Whyte, L. G., Ronholm, J., Prévost, M., and Faucher, S. P. (2020b). Presence of *Legionella* spp. in cooling towers: the role of microbial diversity, pseudomonas, and continuous chlorine application. *Water Res.* 169, 115252. doi: 10.1016/j.watres.2019.115252
- Pereira, A., Silva, A. R., and Melo, L. F. (2021). *Legionella* and biofilms—integrated surveillance to bridge science and real-field demands. *Microorganisms* 9, 1212. doi: 10.3390/microorganisms9061212
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Ren, H., Wang, W., Liu, Y., Liu, S., Lou, L., Cheng, D., et al. (2015). Pyrosequencing analysis of bacterial communities in biofilms from different pipe materials in a city drinking water distribution system of East China. *Appl. Microbiol. Biotechnol.* 99, 10713–10724. doi: 10.1007/s00253-015-6885-6
- Rogosky, A. M., Moak, P. L., and Emmert, E. A. B. (2006). Differential predation by *Bdellovibrio bacteriovorus* 109J. *Curr. Microbiol.* 52, 81–85. doi: 10.1007/s00284-005-0038-6
- Shaheen, M., and Ashbolt, N. J. (2021). Differential bacterial predation by free-living amoebae may result in blooms of *Legionella* in drinking water systems. *Microorganisms* 9, 174. doi: 10.3390/microorganisms9010174
- Thom, C., Smith, C. J., Moore, G., Weir, P., and Ijaz, U. Z. (2022). Microbiomes in drinking water treatment and distribution: a meta-analysis from source to tap. *Water Res.* 212, 118106. doi: 10.1016/j.watres.2022.118106
- Thomas, V., Herrera-Rimann, K., Blanc, D. S., and Greub, G. (2006). Biodiversity of amoebae and amoeba-resisting bacteria in a hospital water network. *Appl. Environ. Microbiol.* 72, 2428–2438. doi: 10.1128/AEM.72.4.2428-2438.2006
- Tomov, A., Kassovsky, V., Chorbadjiiska, L., Tsvetkova, E., Tsanev, N., and Vencheva, Z. (1982). Lytic activity of *Bdellovibrio bacteriovorus* against bacteria of the family legionellaceae. *Zentralbl. Bakteriologie Mikrobiol. Hyg. A* 252, 96–100.
- Vaz-Moreira, I., Nunes, O. C., and Manaia, C. M. (2011). Diversity and antibiotic resistance patterns of sphingomonadaceae isolates from drinking water. *Appl. Environ. Microbiol.* 77, 5697–5706. doi: 10.1128/AEM.00579-11
- Wadowsky, R. M., and Yee, R. B. (1983). Satellite growth of *Legionella pneumophila* with an environmental isolate of *Flavobacterium breve*. *Appl. Environ. Microbiol.* 46, 1447–1449. doi: 10.1128/aem.46.6.1447-1449.1983
- Wang, H., Edwards, M. A., Falkinham, J. O., and Pruden, A. (2013). Probiotic approach to pathogen control in premise plumbing systems? a review. *Environ. Sci. Technol.* 47, 10117–10128. doi: 10.1021/es402455r
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Willis, A., and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics* 71, 1042–1049. doi: 10.1111/biom.12332
- Willis, A., Bunge, J., and Whitman, T. (2017). Improved detection of changes in species richness in high diversity microbial communities. *J. R. Stat. Soc. Ser. C Appl. Stat.* 66, 963–977. doi: 10.1111/rssc.12206