Microbiome in drinking water treatment and distribution: a critical review and meta-analysis from source to tap

Claire Thom[1,2]*, Cindy J Smith[1], Graeme Moore[2], Paul Weir[2]*, Umer Z Ijaz[1]

*Corresponding author

[1]: Infrastructure and Environment Research Division, James Watt School of Engineering, University of Glasgow

[2]: Scottish Water, 6 Castle Drive Dunfermline, KY11 8GG

Highlights

- Meta-analysis of 27 16S rRNA studies from drinking water systems, comprising 1994 samples, identifying 4556 AVS from full length 16S rRNA gene sequences
- We have demonstrated that DWDS microbiomes are more strongly affected by stochastic processes.
- Chlorine has a stronger selective pressure on the assembly of the microbiome than filtration processes
- Pathogens such as *Firmicutes* form co-exclusionary relationships with other phyla after the addition of chlorine.
- *Legionella* abundance may be a good indicator of treatment performance for water utilities.

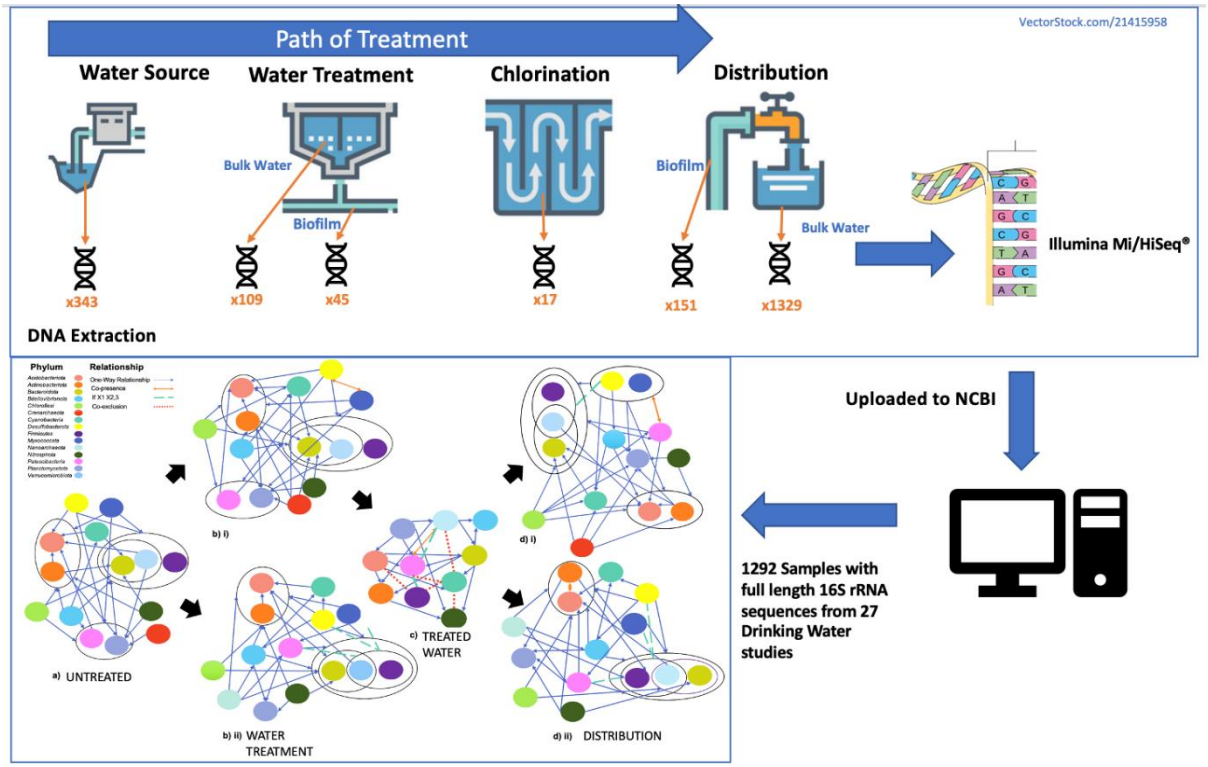
- Changes in species abundance and richness may be useful in detecting contamination in DWDS for water utilities.

Abstract

A meta-analysis of existing available Illumina 16S rRNA datasets from drinking water source, treatment and DWDS were collated to compare changes in abundance and diversity throughout. Samples from bulk water and biofilm were used to assess principles governing microbial community assembly and the value of amplicon sequencing to water utilities. Individual phyla relationships were explored to identify competitive or synergistic factors governing DWDS microbiomes. The relative importance of stochasticity in the assembly of the DWDS microbiome was considered to identify the significance of source and treatment in determining communities in DWDS. Treatment of water significantly reduces overall species abundance and richness, with chlorination of water providing the most impact to individual taxa relationships. The assembly of microbial communities in the bulk water of the source, primary treatment process and DWDS is governed by more stochastic processes, as is the DWDS biofilm. DWDS biofilm is significantly different to bulk water in terms of local contribution to beta diversity and in types of taxa present. Water immediately post chlorination has a more deterministic microbial assembly, highlighting the significance of this process in changing the microbiome although elevated levels of stochasticity in DWDS samples suggest that this may not be the case at customer taps. 16S rRNA sequencing is becoming more routine and may have several uses for water utilities including detection and risk assessment of emerging pathogens like *Legionella*, *Bacteroides* and *Mycobacterium*; assessing the risk of nitrification of DWDS; improved indicators of process performance and monitoring for significant changes in the microbial community to detect contamination.

Combining this with other quantitative methods like flow cytometry will allow a greater depth of understanding of the DWDS microbiome.

Graphical abstract



Key Words

Drinking water, amplicon sequencing, microbiome, pathogens, bacteria, biofilm

1.Introduction

The safety of drinking water supplies is of paramount importance for public health. Water utilities are responsible for the treatment and delivery of potable water. While treatment is highly effective to remove traditional faecal indicator organisms, the microbial challenge remains significant as water-borne disease outbreaks associated with drinking water distribution systems (DWDS) still have significant public health implications, which may not

correlate with traditional water quality metrics[1–5]. Water treatment processes, such as filtration and disinfection, are the primary barriers to the presence of harmful pathogens in drinking water. These are commonly employed strategies for water treatment across the world. Coagulation of colloidal material and subsequent filtration through rapid gravity sand filters, followed by disinfection with a chlorine-based biocide remains the most common method of treatment in many countries. Although there are several other treatment strategies including: slow sand filtration, biological filtration, ozonation and membrane filtration which satisfy drinking water regulations.

Regulations on the microbial safety of drinking water supplies focus on the likelihood of faecal contamination, using presence of coliform bacteria and specifically *E. coli* as a surrogate for the wide range of pathogens potentially present in faeces. These are measured using culture-based tests that isolate and enumerate coliforms and *E. coli* specifically. These methods have been broadly unchanged for over 100 years. Compliance with these metrics is high in the UK (>99%)[6], although isolated, sporadic, and low-level total coliform detections remain a problem for utilities. These indicators are now known to be problematic in that, more than 99% of bacteria are unculturable[7,8]; there are emerging pathogens in drinking water which are not faecal associated, e.g., *Mycobacterium* and *Legionella*; and the correlation between total coliforms and other pathogenic indicators is poor[2,9,10]. Moreover, culture tests assess a small volume (~100 mL), and a confirmed result takes over two days, meaning rapid changes in water quality are unable to be detected. There is therefore a need for alternative high-throughput methods of microbial characterisation to assess the diversity of microbial communities across space and time. Quantitative polymerase chain reaction (qPCR) methods are capable of identification of specific pathogens or target organisms but are limited in the amount of information they give about the overall microbiome. In contrast, Flow Cytometry

(FCM) is a rapid high-throughput methodology to count all intact and damaged bacterial cells *in situ* without the need for culturing. This has been adopted by water utilities in recent years due to its accuracy and speed but can only give an indirect view of water quality risk as cells are not taxonomically identified. It should also be stated that despite its limitations culturing remains the only direct method of determining a cells viability.

The treatment of drinking water in general reduces the abundance and diversity of micro-organisms present, while removing harmful pathogens, yet a diverse microbiome remains associated with potable water, including pathogenic microbes. The drinking water microbiome at customer taps may be influenced by a range of factors including source water, treatment, flow conditions and DWDS biofilms. Water treatment is proposed to have a deterministic effect, selecting microbes that survive filtration and disinfection processes[11,12]. This effect is likely to reduce with distance and time from treatment, where biofilm growth and disturbance become more prominent. At this point stochastic (random) effects are more likely to govern the assembly of microbial communities. Thus, drinking water microbiomes are dynamic through treatment, time, and location. To aid water utilities direct and control these to deliver safe potable water a deeper understanding of these changes, consequences, and impact on both the microbiome and the prevalence of pathogens is needed, moving from descriptive to predictive understanding.

16S rRNA amplicon sequencing technology can be used to characterise and identify microbial communities in DWDS across space and time. While the taxonomic resolution that can be achieved depends on the 16S rRNA hypervariable region sequenced and the type and abundance of the taxa detected[13], the approach has been widely applied in academia since circa 2010 to explore microbial communities of drinking water to both assess diversity and

identify pathogens of concern to public health. This method has broad advantages over qPCR and FCM in the large amount of taxonomic information that it provides, although its disadvantages are that it is non-quantitative and cell viability is unable to be determined. Most 16S rRNA studies are discrete, commonly across a single or few DWDS within a geographical area. Studies tend to focus on one part of a system, e.g., source waters; efficacy of treatment processes; variations of biofilm communities in space, time, or operating conditions within a pipe distribution network; influences of domestic plumbing arrangements, or differences between the bulk water and biofilm communities[14–43]. While several of these studies have provided new insight into drinking water microbiome, as aforementioned they tend to be descriptive and not predictive. There is a need for further insight, at a global level, into the principles governing drinking water microbiomes to gain further understanding of the ecological rules determining microbial assembly through the treatment process to the tap if amplicon sequencing is to be of use to utilities. Here we present a critical review of current understanding and further conduct a meta-analysis of 16S rRNA studies from source to tap to explore global distribution and commonalities in the drinking water microbiome. We further consider the contribution and potential of 16S rRNA amplicon sequencing as an analytical tool for water utilities as a common conclusion to several studies is that 16S rRNA amplicon sequencing is beneficial in assessing risk to public health in DWDS, although there are many areas for further investigation to understand the implications of the results[14].

Source Waters

Several studies have sought to assess the impact of source water on the microbial communities in DWDS, with groundwater sources less phylogenetically diverse than that of surface waters[15]. Some studies propose that the composition of taxa in DWDS are strongly determined by those which are present in the source water[15]. For example, land-use in the

catchment of source waters can impact the microbial communities present in terms of abundance and diversity, as demonstrated by significant differences between urban and agricultural catchments in Canada. *Bacteroides* was identified as a potential future indicator of source contamination for both catchment types[37]. 16S rRNA amplicon sequencing combined with microbial source tracking can identify contamination sources, although these methods correlate poorly with each other and traditional faecal indicators[38]. Eighty-one potentially pathogenic bacteria were detected within drinking water source waters using 16S rRNA amplicon sequencing, with little correlation to indicator organisms like coliforms and *E. coli*[26]. These studies show that amplicon sequencing has the capability to detect pathogenic microbes in drinking water systems, although further assessment of how this relates to regulatory parameters like indicator bacteria is required. It also must be noted that the presence of DNA does not necessarily imply viability, with DNA from dead cells and extracellular DNA contributing to DNA extractions. Prior sample treatment, for example with propidium monoazide[44,45], may aid inform viability by removing cell with damaged cell walls prior to DNA extraction. Yet further investigation would still be required to confirm cell viability, a current limitation of 16S rRNA approaches.

Water Treatment

Some individual studies have focused on the relative importance of treatment in influencing downstream microbial communities, with the processes of filtration and chlorination reducing overall microbial diversity and affecting the abundance of particular phyla[39–43]. Biofilters can be extremely microbially diverse with different operational conditions causing changes to relative taxa abundances[39,43,46]. The upstream process also significantly affects filter biomass[16,17]. Various treatment and disinfectant types have variable effects on the abundance of different phyla, with no-one treatment type effectively removing all pathogens[18,19].

Although chlorination effectively reduces microbial abundance, pathogens such as *Pseudomonas, Acinetobacter, Citrobacter, Mycobacterium, Salmonella, Staphylococcus, Legionella, Streptococcus,* and *Enterococcus* have all been detected in treated water, indicating that further work is needed to assess and improve the robustness of treatment[42]. Overall, the microbial diversity in individual published studies were highly specific to the treatment process employed and a range of study specific pathogens were detected. Therefore, wider assessment of treatment and DWDS is required to understand whether these pathogens are specific to those systems or a wider problem in drinking water. In addition, organisms responsible for nitrification such as *Nitrospira* are also seen to be abundant throughout treatment and into DWDS[20–22].

Drinking Water Distribution Systems (DWDS)

The majority of 16S rRNA amplicon sequencing studies have focused on microbial communities within DWDS, in both biofilm and bulk water. These indicate that the microbial communities within DWDS are unique to that system, influenced by the source and treatment characteristics outlined above[23,24]. DWDS microbial communities have significant temporal fluctuation[25,27], with potential diurnal cycles in bulk-water, potentially due to flow patterns[28]. The dynamics of microbial communities in bulk water may be seasonal. Higher diversities and abundances of microbial communities may be more evident in winter compared to summer months[19]. DWDS may also have significant spatial variation[19]. A meta-analysis of 14 pyrosequencing studies of water distribution systems compared the bacterial communities present under different disinfectant regimes, confirming that the microbial communities in DWDS are more diverse and abundant than those with a free chlorine residual. *Legionella, Mycobacterium,* and *Pseudomonas* were all significantly reduced by the presence of a free chlorine residual in one study[28]. However, it is unknown whether free chlorine residual is

significant in reducing diversity across all DWDS or whether source and treatment processes may also play a significant role. 16S rRNA amplicon sequencing is a powerful tool to aid identify risk, as it can identify potential pathogenic microbes in distribution[21,29–32]. However, as mentioned above, it is unable to distinguish between alive or dead cells. What is unclear from the individual studies is whether certain pathogens are common to DWDS in general, and what is their source.

Biofilm vs. Bulk- Water

Biofilm and bulk water samples in distribution have also been shown in several individual studies to vary significantly[30,33–35]. Biofilm deposition influences bulk water communities when loose deposits or biofilm are disturbed[30,35,47]. Biofilms can significantly contribute to microbial loading in DWDS, with the composition affected by the presence of a chlorine residual[36]. Mechanical cleaning also changes the microbial composition of biofilms, with a lag effect in bacterial concentrations observed after cleaning followed by a regrowth phase[36]. A recent study in Sweden explored how the bulk water and biofilm in DWDS were affected by ultrafiltration membrane (UF) installation using source tracking software. Bacteria in the bulk water came from treatment (99.5%) before the installation of UF. Post-UF, there was a significant reduction in cells, and 58% were quantified to have come from biofilm in the DWDS[22]. These results suggest that when large volumes of bacteria are removed throughout treatment processes, the relative influence of the biofilm may become larger. As biofilms have been demonstrated to differ from bulk water in terms of taxa and potentially harbour pathogens, this is likely to affect water quality. Further understanding of how treatment, source, and biofilm in distribution affect DWDS microbiomes is required for water utilities to understand the impacts of changing treatment processes.

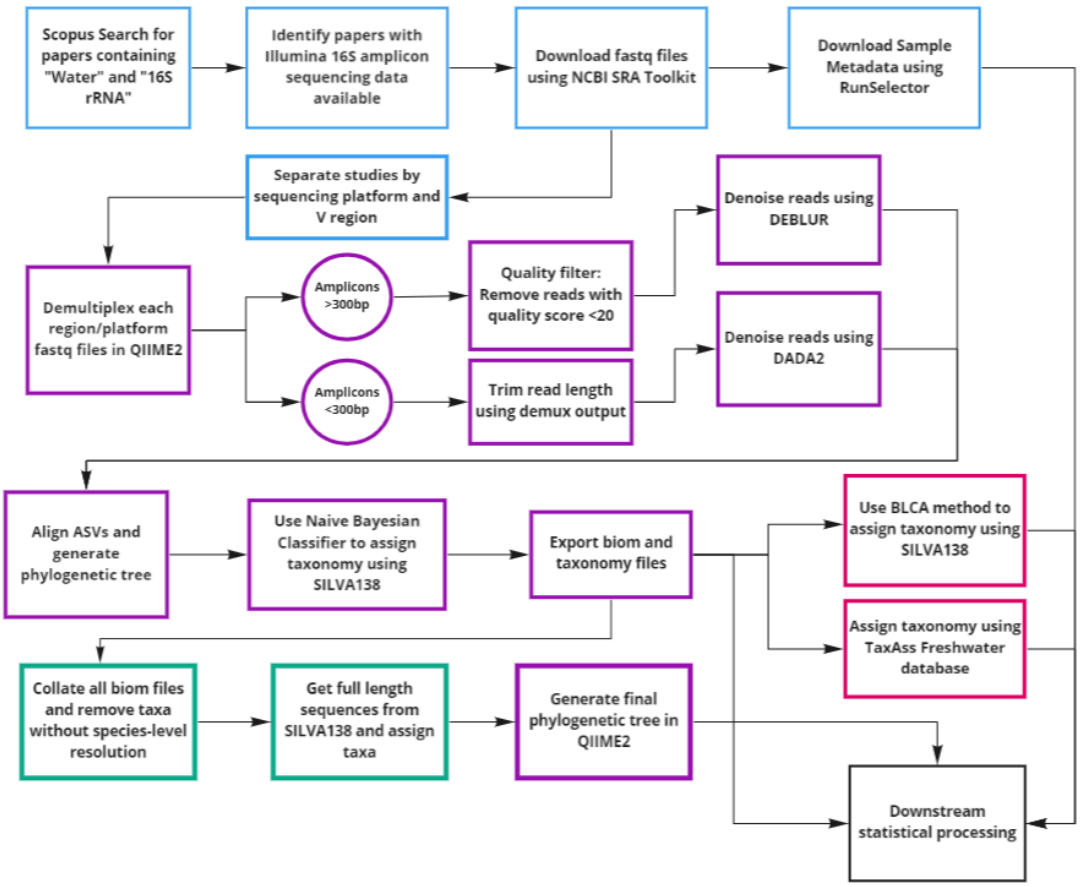
Aims and objectives

While these studies have provided new understanding of drinking water microbiomes a limitation is that they are in general descriptive snap shots of a given time/treatment etc that by themselves limit current ability to define global rules to facilitate predictive management of water treatment and quality. Add to this variability in the methods used (DNA extraction, 16S rRNA hypervariable region, sequencing platform etc) and the complex system-specific nature of DWDS, all contribute towards making it difficult to gain a more general understanding of how drinking water microbial communities change from source to tap and what are the factors driving these changes that can be utilised by water utilities to manage water quality. For example, many studies have identified pathogens of concern in individual systems, but it is still unclear whether these pathogens are common to all DWDS, or an artifact of specific systems. The relative importance of the deterministic influence of source and treatment is also an important question as is the need to understand how phyla common to all DWDS interact with each other as competitive influences between taxa may affect the microbiome. The lack of understanding of the principles governing the diversity and abundance of DWDS microbial communities has made it difficult to predict/control drinking water microbiomes and therefore also for water utilities to adopt the understanding derived from amplicon sequencing approaches as a tool to aid manage water quality, including pathogen detection.

Therefore, the specific aims of this meta-analysis were to identify commonalities in DWDS microbiomes across the world, which can be used to further understanding of water quality for utilities; to understand the relative importance of the deterministic effects of source and

233 treatment on the microbiomes of DWDS and explore key relationships between phyla
234 present.

235 2.Methods



236
237
238 Figure 1: An overview of the methodologies to generate Amplicon Sequencing Variants (ASVs) applied in this meta-
239 analysis.
240

241 2.1 Data Gathering

242 A Scopus search for all papers since 2010 using the following terms: “16S rRNA” and
243 “Water” was carried out. This search returned 176 results. Each result was individually
244 assessed to ascertain its relevance to this meta-analysis. Only studies using Illumina MiSeq or
245 HiSeq® platforms were included to minimise the different errors and biases associated with

alternative sequencing platforms such as Nanopore®, Ion Torrent® or older technologies such as Pyrosequencing. After this manual filter, 44 studies remained and were checked to ascertain sequence data availability. 26 studies had publicly available raw sequence data. For the remainder, requests to authors were made for data. A list of the papers used in the analysis can be found in the supplementary information. All raw data downloads used the SRA Toolkit provided by NCBI, except for one study from QIITA. Metadata for samples from NCBI's Run Selector included: sequencing platform; the hypervariable region of the 16S rRNA gene sequenced; sample ID; sample date and time; and geolocation. Other relevant metadata from the published papers: sample location, disinfection type (if applicable), and whether the sample was from bulk water or biofilm was recorded. Before processing, studies were grouped by the hypervariable region of the 16S rRNA gene sequence. All studies included in this meta-analysis and relevant sample information are listed in the Supplementary Information section. In total 27 studies, with 1994 samples, from over 50 different DWDS were compared.

2.2 Sequence Processing

QIIME2 processed collated amplicon sequences for each platform and hypervariable V-region in Earth Microbiome Project Paired-end Sequencing Format (.fastq). QIIME2 can generate both Operational Taxonomic Units (OTUs) and Amplicon Sequencing Variants (ASVs) using a user-defined threshold (97% in this case). QIIME2 improves QIIME1 in terms of quality control of sequences using DADA2 and Deblur software, both of which were employed here. To provide enough overlap of forward and reverse reads to facilitate paired end reads, DADA2 was employed where amplicons were <250bp long and the quality score was >20. For amplicons spanning multiple V regions, DEBLUR commands allowed for the pairing of longer amplicons without significant loss of sequence length, as an explicit

threshold is not required. Output alpha diversity profiles may be significantly different when using different denoising software to generate ASVs[48], so, runs of DEBLUR and DADA2 were carried out for all regions and platforms. The final analyses generated 3.32 X 10^8 demultiplexed reads and 829713 ASVs in total from 1994 samples.

To identify the best taxonomic assignment, biome files and phylogenetic tree output from QIIME2 had taxonomy assigned using three approaches. These were: Naïve Bayesian Classification system (NBC), Bayesian Least Common Ancestor (BLCA) approach (using SILVA138 database), and the TaxAss database. TaxAss uses SILVA to generate a first pass of taxonomic assignment then a curated database of freshwater sequences to assign the remainder of ASVs. TaxAss was selected for downstream statistical processing as it provided the highest level of taxonomic recovery to the genus level (Appendix 1 Table 1). Finally, ASVs from all V regions were collated together in a single biome file. Sequences without species-level resolution were removed so that full-length 16S rRNA sequences could be obtained for all taxa as per the method used by[49] Of the 4858 taxa originally classified by TaxAss in the collated dataset, 4556 had available full-length sequences (loss of 6.2%). A final phylogenetic tree and biome file with taxonomy generated in QIIME2.

2.3 Statistical Analyses

The collated biome with taxonomy, phylogenetic tree, and metadata was then processed. Meta-sample groupings defined the sample location in the treatment and distribution process and if the sample originated from biofilm or bulk water. Shannon and Richness indexes were calculated for each meta-grouping to estimate alpha diversity (diversity within a sample). An analysis of the core microbiome was carried out in the R package Bioconductor, using an absolute detection method and a minimum prevalence of 85% for all groups except

296 Distributed and Untreated Water. These groups had significantly more samples and required
297 a higher threshold of 95% (Lahti *et al*. 2017-2020[50]).
298
299 Beta diversity (or between-sample diversity) metrics were more complicated to assess, given
300 the substantial number of samples (n=1994), their varying environments as well as spatial
301 and temporal locations. Instead, calculation of Local Contribution to Beta Diversity (LCBD)
302 for each group was made[51]. The Nearest Taxon Index (NTI) and Net Relatedness Index (NRI)
303 from the Picante package in R (http://kembellab.ca/r-
304 workshop/biodivR/SK_Biodiversity_R.html) were used to quantify Environmental filtering
305 and stochasticity on community assembly. Higher NRI/NTI (above 0) values indicate
306 deterministic factors influencing community assembly, lower values (below 0) indicate
307 stochastic influences.
308
309 Patterns in beta diversity may not be continual, as multiple relationships may be affecting an
310 organism at a specific time or place. Therefore, a new methodology by Golovko *et al.* (2020)
311 was employed using Boolean patterns to assess relationships between individual ASVs in all
312 meta-sample groups. This method uses a pattern-specific method to identify 2-dimensional
313 relationships between 2 ASVs at a defined threshold, including one-way relationships, co-
314 occurrence, and co-exclusion. This method can also quantify 3-dimensional relationships
315 between ASVs. These are categorised as: all alone (type 1 co-exclusion); exclusion of ASV1
316 by ASV2 or 3 (type 2 co-exclusion); If ASV1 is present, ASV2 and ASV3 are present, and
317 finally, all three altogether. This method was applied to the ASVs in the dataset to identify
318 any significant relationships at a phyla level.
319

3.Results

3.1 Taxonomic Profile

Within this analysis, taxonomic classification was resolved for a total of 4556 ASVs. The 25 most abundant genera are shown in Figure 1. 1293 samples were from water distribution systems in bulk water, the largest meta-sample group. Bulk water from different distribution systems, as expected, is variable with differences in the abundances of the top 25 genera DWDS. However, there does appear to be some commonalities in taxa among DWDS with the same disinfectant residual: *Nitrosomonas and Pseudomonas* were abundant only in systems using a chloraminated residual. Pathogenic microbes such as *Mycobacterium* were common in both chlorinated and chloraminated systems. Biofilm samples in distribution were less numerous (n=193) and had a much higher taxonomic diversity than the bulk water. *Pseudomonas* was common in many samples in both chlorinated and chloraminated biofilms, but less so in those with no disinfectant residual.

Samples from water sources and treatment systems made up a much smaller proportion of the dataset and had differences in the most abundant taxa. Again, the most common genera were less abundant, except *Nitrospira,* which was more abundant throughout treatment than in distribution. *Burkholderiales* were also present throughout treatment and highly abundant in one bulk water study in the distribution. Source water samples were generally from surface waters, although a small proportion came from groundwater. Several untreated water samples show similar taxonomic profiles to each other. Globally all untreated water samples were highly diverse in comparison to treated water.

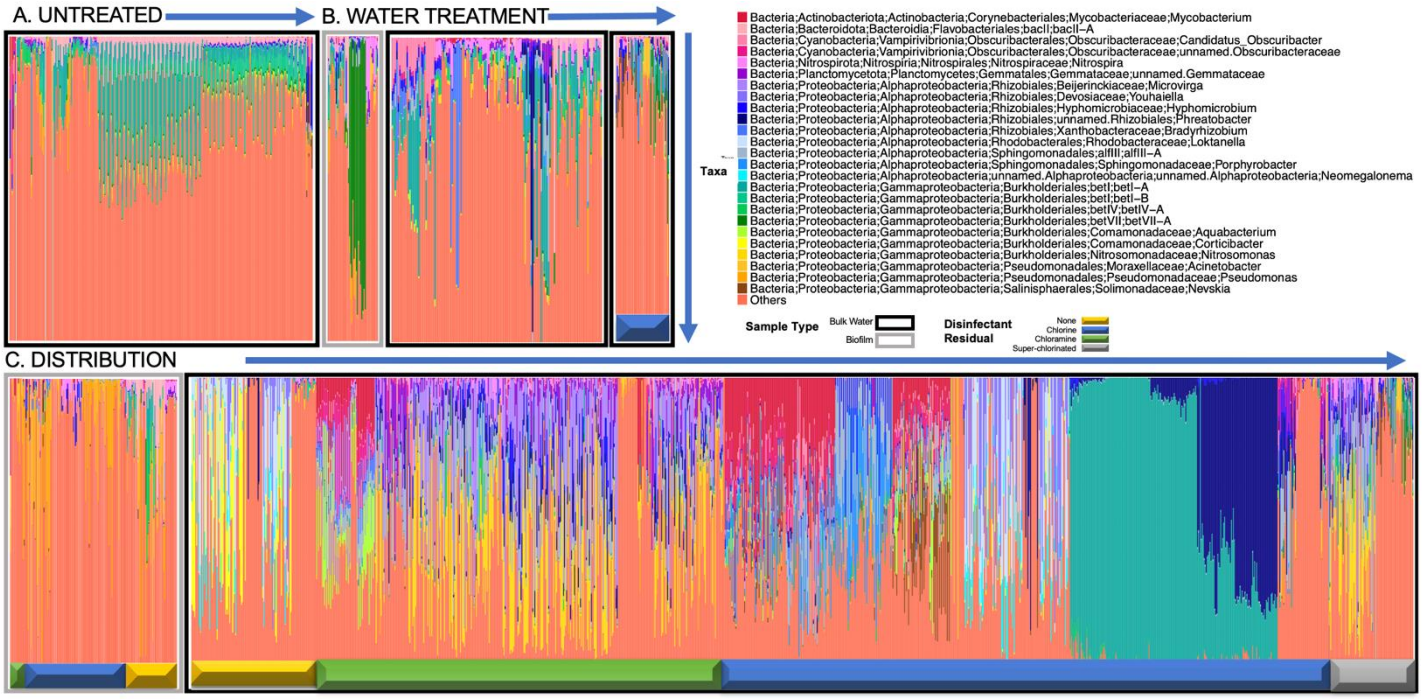


Figure 2: Proportion of most abundant genera for all samples in this meta-analysis, grouped by: A untreated water; B water and biofilm isolated throughout treatment processes; C biofilm and bulk water sampled from distribution pipes. Where appropriate, samples have been colour-coded by disinfectant residual: Free chlorine (blue); Chloraminated (green); None (yellow); Super-chlorinated (grey).

3.2 Alpha Diversity & Core Microbiome

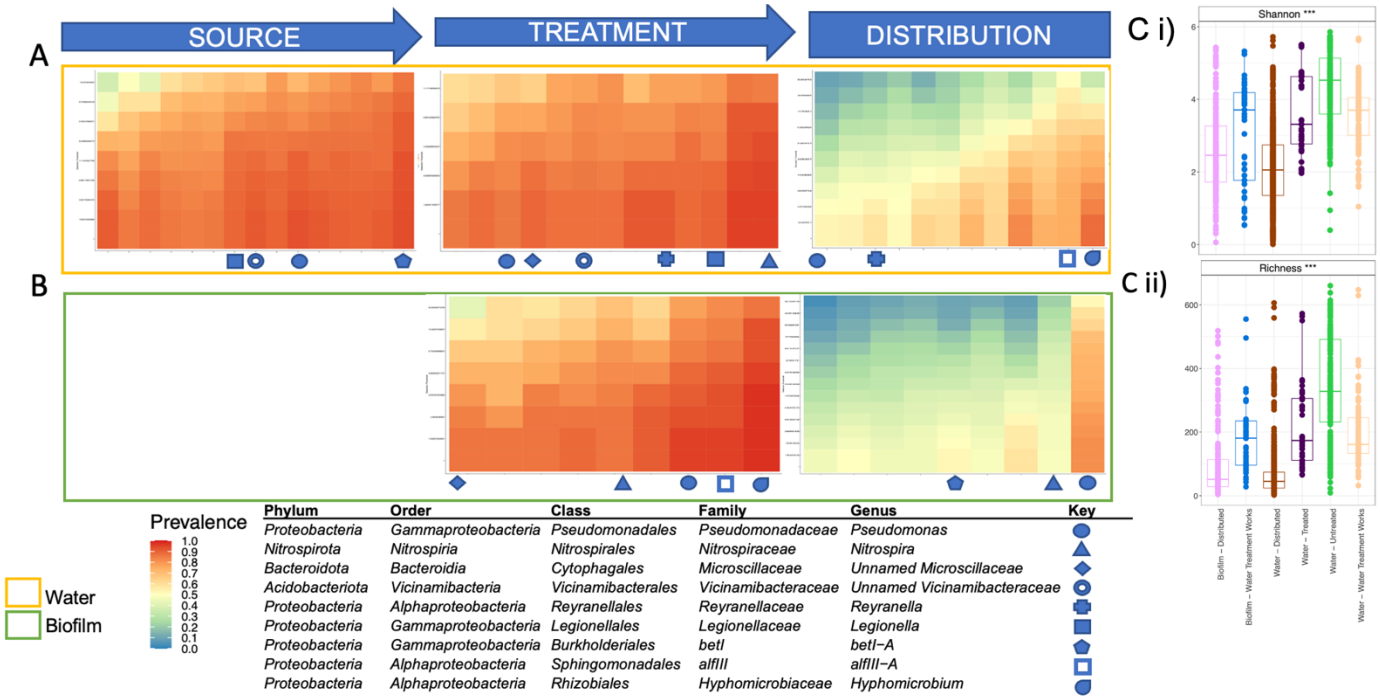


Figure 3: A. Core microbiome analysis of the different meta-sample groups from source through treatment and distribution for bulk water and B for biofilm. Minimum prevalence was set at 0.85 for all groups except Distributed Bulk Water and Untreated water, set at 0.95 due to the high number of samples in those groups. C: the alpha diversity of the various meta-sample groups, displaying i) Shannon values and ii) Richness.

The amount of diversity within each sample, or alpha diversity, can be seen in Figure 3(C). Across the different sample groups, the within-sample Richness values were significantly different. The highest degree of sequence diversity in terms of Richness and Shannon index values came from untreated water. A reduction in these values was evident in the treatment and distribution groups, in biofilm and bulk water. Biofilm samples have elevated Shannon values compared to bulk water, although Richness was remarkably similar. Core microbiome analysis proposed several prevalent taxa within more than one sample group, although the overall taxa prevalence reduced in distribution samples. *Pseudomonas* was the only taxa common to all stages of water treatment and distribution. *Nitrospira* was prevalent within water treatment works and in distribution biofilm, but not bulk water. *Legionella* was

abundant in bulk water only, in untreated and in treatment samples. *Burkholderiales betI-A* was prevalent in source water and distribution biofilm samples.

3.3 Local Contribution to Beta Diversity

Due to the unequal data classes with high degrees of spatial and temporal variation, estimates of beta diversity used Local Contribution to Beta diversity (LCBD) for all meta-sample groups (Figure 4) rather than a direct measure of beta diversity. LCBD values were only above the significance threshold for two categories when calculated using Unifrac distance: untreated water and distribution biofilm. For Bray-Curtis, all groups had greater than the calculated threshold (0.00052) LCBD except distribution water, with biofilm samples having the highest value. NTI and NRI values for the meta-sample groups were similar except for untreated water, which was the only category with values <0, indicating the taxa present are more dissimilar than in the other categories. Biofilm and bulk water in distribution had almost equal NRI/NTI indicating no significant difference in the amount of species relatedness between these groups.

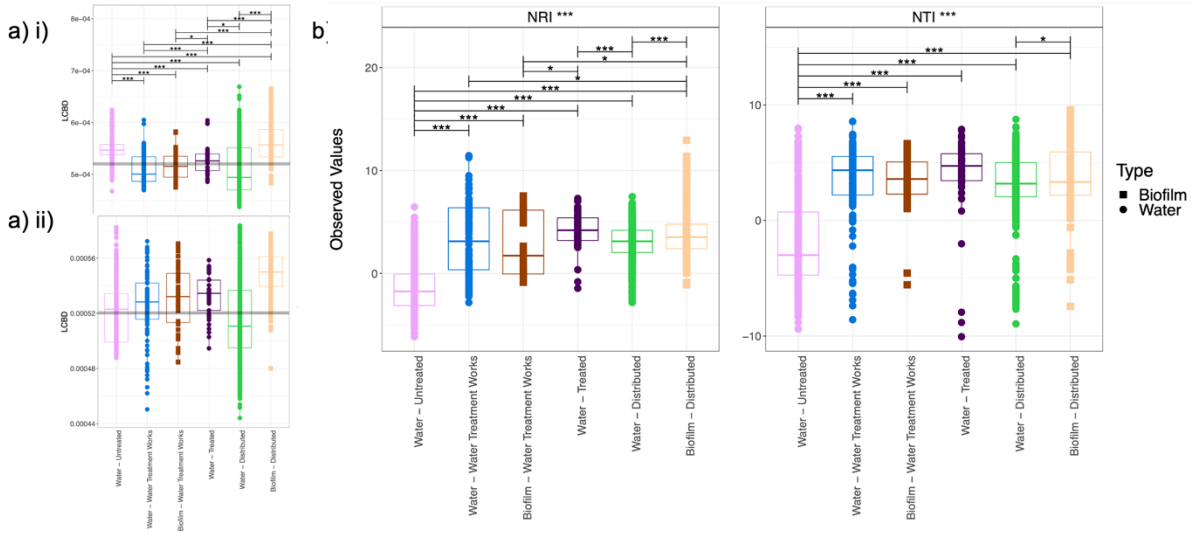


Figure 4: a) Local contribution to beta diversity values for all meta-sample groups using Unifrac (i)) and Bray (ii)). B) net relatedness index (NRI) and nearest taxon index (NTI) of all meta-sample groups in this meta-analysis.

3.4 Normalised Stochasticity Ratio

The NST and MST displayed in Figure 5 quantify the relative importance of stochasticity for each meta-sample group. Phylogenetic distances calculated using Jaccard with and without abundances (Ruzicka approach). Both measures produced comparable results. NST values of >0.5 are considered to be more stochastic. Untreated and samples prior to disinfection had NST values greater than 0.5, as did those in distributed bulk water. Meta-samples immediately post disinfection (treated water) had reduced stochasticity (0.45), indicating a greater degree of determinism in community assembly at this stage of the process. Biofilms in water treatment works had the lowest NST values, 0.42, showing higher determinism within these samples. MST values (modified ratio) are much lower, although the most deterministic groups are the same as in the NST. Untreated water samples have the highest degree of stochasticity when using MST compared to distributed bulk water and biofilm samples when using NST.

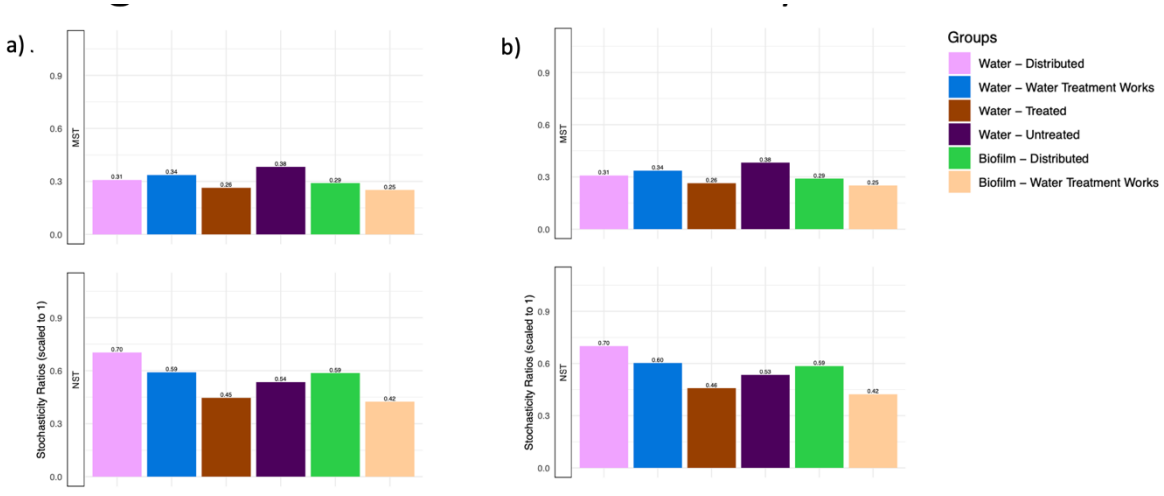


Figure 5: Normalised Stochasticity ratio (NST) and b) Modified Stochasticity Ratio MST) values for all meta-sample groups using a): Jaccard measures of phylogenetic distance and b) Ruzicka measures. Ruzicka is as Jaccard except those relative abundances are not considered.

3.5 Boolean Relationships

The results of the Boolean analysis to identify individual relationships between ASVs in the dataset at 2 and 3-dimensional levels is displayed in Figure 6.

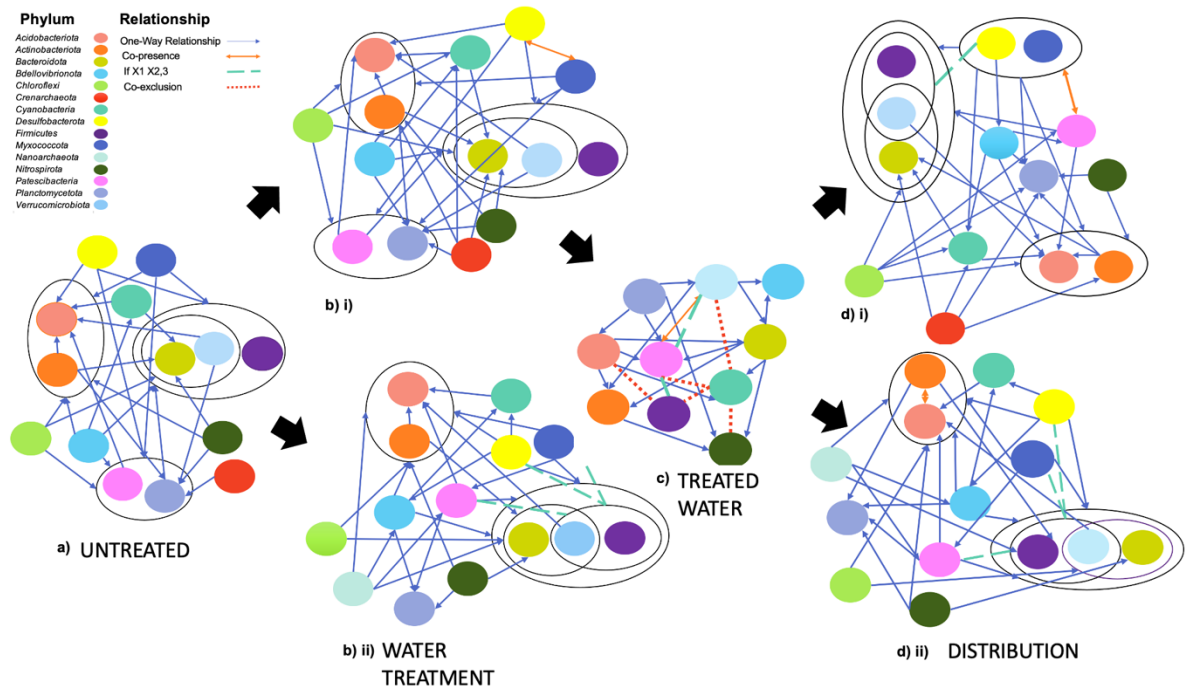


Figure 6: Visual representation of individual Boolean relationships identified between Phyla for all meta-sample groups in this study, using methodology from Golovko et al. (2020). Minimal presence threshold set at 0.05%, maximum at 0.1%. Groups are a) Untreated water; b) i) samples taken from bulk water throughout treatment; b) ii) biofilm from water treatment processes; c) disinfected water; d) i) water from distribution systems and d) ii) biofilm from pipes in distribution. 2-dimensional relationships detected included one-way relationships (blue arrow); co-presence (orange arrow), and co-exclusion (red dashed line). There was one 3-dimensional relationship if ASV 1 is present, ASV 2 and 3 are also present (long dashed line).

The analysis identified many one-way relationships between individual phyla across all stages of the treatment and distribution process, many of which were present in several

groups. These phyla were common to all sample groups except *Nanoarchaeota* and *Crenoarchaeota*, present only in distribution bulk water and biofilm, respectively. Many of these phyla exhibited this relationship in all groups, such as *Acidobacteriota* and *Actinobacteriota*. When *Acidobacteriota* is present, *Actinobacteriota* is also present. This relationship is one-way in all categories except biofilm in the distribution system, where they exhibit a co-presence relationship. Treated water samples had the lowest number of phyla relationships within the dataset, and this was the only meta-group to have any co-exclusion relationships detected. *Cyanobacteria* were in a co-exclusionary relationship with three other phyla: *Verrucomicrobiota*, *Firmicutes*, *Nitrospirota, and Patescibacteria*. *Acidobacteriota* and *Firmicutes* were also co-excluded. Relationships from samples in distribution were more like those in the untreated and partially treated samples, with large numbers of one-way relationships between the Phyla in biofilm and bulk water.

4.Discussion

4.1 Principles governing microbiomes in DWDS

Taxonomic profiles of the various meta-sample groups identified significant differences in abundances of genera throughout the source, treatment, and distribution of water. There are higher degrees of species richness and alpha diversity of taxa in the source waters that are reduced throughout treatment. This reduction is consistent with both the individual studies included in this analysis and several other pyrosequencing studies[23,31,52]. The reduction in LCBD from untreated to treated water samples supports the reduction in alpha diversity and richness. A wide-scale study of 49 distribution systems in China had reduced diversity and

Richness in tap compared to source waters[15]. The similarity of taxa increases throughout treatment and distribution, indicative of selective processes driven by filtration and chlorination from which only some organisms can survive[39–43].

Biofilm and bulk water samples from DWDS are quite different in terms of types and abundance of taxa. The core microbiomes for both these groups had only *Pseudomonas* common to both. Although Pseudomonas was the most abundant organism within biofilm core microbiome, it was the least abundant in bulk water samples. Local contribution to beta diversity was also most elevated in biofilm samples and was the only sample above the calculated threshold. This indicates that biofilm microbiomes contribute more to overall biodiversity within the pipe than the bulk water samples. This is important to consider, as the biofilm contains a quite different microbial profile than that of the bulk water, and sampling only bulk water may give a limited picture of the overall microbiome.

Modelling of microbiomes concentrates on the relative importance of random events on assembly, such as births, deaths, and environmental disturbance, this is compared to more deterministic events such as selection. As treatment and sources may be significant in determining the organisms in the treated water, this is an important measure to consider. All meta-sample groups, except treated water and distributed biofilm, had more stochastic values suggesting a greater degree of randomness in microbiome assembly. However, samples immediately after disinfection with chlorine had a more deterministic value. *Proteobacteria*, in particular: *Pseudomonas*, *Actinobacter*, and *Rheinheimera* have been demonstrated to dominate post disinfection, supporting the deterministic influence of treatment[53]. It also suggests that although filtration is important in defining taxa in DWDS, chlorine has a more strongly selective effect. This hypothesis supports a study comparing two identical treatment

474 systems treating the same source water, where chlorine and chloramine produced different
475 bacterial communities[23]. Biofilms in water treatment works also had higher NST values,
476 indicating a higher degree of determinism. The higher value supports the hypothesis that the
477 effects of treatment and source reduce with distance and time from treatment[15,19,52]. This
478 increased determinism could be due to the influence of prior treatment and source water on
479 the biofilm, but also the material and conditions within the pipe. There may be other factors
480 too, such as the flow conditions, within the DWDS influencing the biofilm. This had been
481 demonstrated by laboratory experiments using experimental pipe loops with the same influent
482 water under different flow rates resulting in biofilms containing some shared core taxa but
483 with differences in their relative abundances, influenced by the flow conditions within each
484 loop[30].
485
486 Identification of significant one, two, and three-way relationships between individual Phyla
487 in the meta-sample groups demonstrates the complexity of the drinking water microbiome.
488 Several Phyla are proposed to be important in DWDS. A long-term study of a drinking water
489 microbiome identified seven dominant phyla (*Proteobacteria, Bacteroidetes, Actinobacteria,*
490 *Nitrospira, OD1, Planctomycetes, and Acidobacteria*)[12]. Six further phyla were also proposed
491 by an Ion torrent study of DWDS in the Netherlands (*Chloroflexi, Elusimicrobia,*
492 *Chlamydiae, Firmicutes, TM7, and Verrucomicrobia*)[54]. A number of these Phyla are
493 involved in at least one one-way relationship with other phyla within all meta-sample
494 groups: *Actinobacteria; Bacteriodetes; Planctomycetes; Chloroflexi; Firmicutes;*
495 *Verrucomicrobiota*. This further supports their importance in the microbiome of DWDS.
496 Understanding these relationships may be a crucial first step in shaping a biostable
497 microbiome throughout the DWDS which in turn could be exploited to control pathogens.
498 For example, *Firmicutes,* which may contain several known pathogenic organisms, form co-

499 exclusionary relationships with other phyla, post disinfection with chlorine, suggesting
500 competition between them and *Cyanobacteria*, *Acidobacterio*ta, and *Patescibacteriota*. The
501 Boolean analysis confirmed that overall, phyla relationships are reduced by the addition of
502 chlorine. In the future can this new understanding as shown here be exploited to shape
503 microbial communities to the exclusion of pathogens.
504
505 This key finding was revealed from a limited comparison of studies primarily focused on the
506 DWDS bulk water. Therefore, there is a need to explore further the influences of source and
507 treatment processes on treatment, DWDS bulk water and biofilm to explore further the
508 ecological rules and relationships between taxa. Further sampling of these stages is required
509 for water authorities to understand the drivers of overall diversity, implication of treatment
510 and distributions and the influence on pathogenic microbes within DWDS.
511

512 ## 4.2 Applying 16S rRNA Amplicon Sequencing for Water Utilities

513
514 As expected, the 25 most abundant taxa in the analysis did not contain any organisms
515 traditionally used to indicate contamination, as these should be in low abundance in treated
516 water. There were no coliforms identified in the core microbiome for any meta-sample
517 groups. There was also a lack of any *Enterobacteria* in the most abundant taxa for any meta-
518 groups' taxa profile or core microbiome analysis, demonstrating that faecal organisms are in
519 low abundance in DWDS. An analysis of untreated water samples in isolation also failed to
520 identify any highly abundant *Enterobacteriaceae.* These results suggest that 16S rRNA
521 amplicon sequencing is not appropriate for the detection of traditional indicator organisms
522 like coliforms but does suggest that these organisms are not a part of the microbiome under
523 normal operating conditions, although whether their detection is indicative of contamination

524 is not covered by this study. Coliform bacteria are considered indicators of process
525 performance, rather than faecal contamination (except *E. coli*) due to their prevalence in
526 some environments and lack of correlation to other enteric pathogens[9,55,56]. This study further
527 suggests that their overall lack of abundance in untreated water makes them a poor indicator
528 of process performance.
529
530 This analysis did reveal some organisms of concern as abundant in DWDS, although
531 different organisms were of concern in different DWDS, consistent with the proposed
532 system-specific nature of DWDS microbiomes[54]. Of note, *Mycobacterium* was abundant in
533 both chlorinated and chloraminated DWDS, but was not prevalent in non-chlorine distributed
534 water. *Mycobacterium* is an emerging pathogen of concern for water utilities and dominates
535 in some DWDS[32,57]. *Nitrosomonas* and *Nitrospira* were also highly prevalent in the biofilm
536 of chloraminated DWDS, supporting the results of individual studies[21,22]. Improving
537 understanding of the processes that select for and remove nitrifiers
538 like *Nitrosomonas* and *Nitrospira,* is also important to assess the risk of nitrification within
539 DWDS. *Burkholderiales* was a highly prevalent member of the core microbiome, consistent
540 with other results proposing that this organism is under selection by treatment processes and
541 is resistant to chlorination[21,58]. Although the high abundance of these organisms was only in
542 one DWDS in this analysis, it highlights the need for more studies to be added to understand
543 if this is a general finding.
544
545 If water utilities can optimise processes to select for non-pathogenic microbes, this can
546 reduce the risk of illness from drinking water, something suggested in several studies[30,35,36].
547 An overview of microbial communities' dynamics, as ascertained by 16S rRNA amplicon
548 studies aids inform a holistic view and response of water treatment and DWDS system

549 microbiology. This understanding will aid management to maintain water quality and will
550 enable control of the drinking water microbiome and pathogens.
551
552 16S rRNA studies are becoming more popular and routine for the molecular analysis of water
553 treatment and distribution. As demonstrated here, they provide extensive information
554 revealing diverse communities that are influenced by the treatment process. However,
555 translating this information into practice to inform and predict water quality is not always
556 obvious to water utilities. However, taking a global meta-analysis view, this analysis
557 highlighted several ways in which water utilities might employ 16S rRNA sequencing to
558 improve drinking water quality. Considering whole microbial community dynamics from
559 source to water, bulk and biofilm, this review has identified several organisms highly
560 abundant throughout source and treatment, that can be potentially used to benchmark
561 performance and monitor risk. *Pseudomonas* and *Mycobacterium* were all abundant in
562 DWDS, while *Legionella* was abundant in source and treatment stages. Members of these
563 group are known pathogens of concern for drinking water quality. In particular, the higher
564 abundance of *Legionella* in source waters and treatment in this analysis may make it a good
565 indicator of treatment performance, especially as other studies have detected *Legionella* in
566 treated water samples[41,42,57]. *Legionella* is an emerging pathogen of concern to the water
567 industry, and in the UK may be included in future water quality regulations. Amplicon
568 sequencing can also allow water utilities to assess the risk of non-compliance with these
569 regulations, although the viability of the organisms must also be considered using an
570 alternative method.
571
572 Flow Cytometry (FCM) may provide the appropriate information to compliment 16S rRNA
573 sequencing. Using FCM with the intercalating dyes SYBr Green and propidium iodide to

574 stain genetic material *in situ* within a sample gives a quantitative measure of the intact cells
575 within the microbiome of DWDS. This is due to the ability of SYBr green to permeate intact
576 cell membranes, whereas propidium iodide cannot. Dead cells therefore appear red, where
577 intact cells fluoresce green, allowing each to be distinguished by the FCM. This approach has
578 been extensively explored in studies assessing water treatment cell removal, DWDS regrowth
579 and seasonal changes within microbiomes[59–62]. FCM can also provide more information than
580 just the count of cells within a sample, using the relative fluorescence and a statistical binning
581 process, cells can be grouped into populations which can then be tracked[63,64]. A quantitative
582 measure of the intact cell population in relation to the total count of cells within a sample
583 could be used by utilities to quantify the viability of the organisms identified using a 16S
584 rRNA amplicon sequencing, enhancing the benefits of both analyses.
585
586 Measures of species richness and abundance such as alpha diversity and LCBD at treatment
587 and distribution stages are useful to water utilities when comparing DWDS performance.
588 Although monitoring the relative abundance of specific taxa in a single DWDS may not be
589 able to detect a risk to public health directly, an understanding of these values across different
590 DWDS allow water utilities to assess the impacts of source, treatment and distribution
591 conditions on water quality and make more informed choices on asset investment.
592 Understanding the relative impacts of stochasticity in DWDS microbiomes is also a useful
593 exercise for water utilities. Higher stochastic values in bulk water and biofilm of DWDS
594 suggest that random events are more important than treatment processes or other prior
595 deterministic events in determining the bacterial communities. In contrast, biofilm
596 communities in water treatment works and bulk treated water are more deterministic, affected
597 by the abiotic conditions (e.g., chlorine, pH, pipe material) and prior treatment processes.

598 Managing biofilm and ensuring treatment processes remove microbes of concern is where
599 water utilities can most effectively minimise risk to public health.

600 5. Conclusions

601 There are copious quantities of data from amplicon sequencing studies in drinking water
602 treatment and distribution. Although many of these studies provide only a descriptive
603 understanding of the microbiome. As a result, this information has yet to be used to predict
604 and direct microbial water quality. There has been a reluctance to adopt the technology
605 among water utilities, as the benefits are not immediately clear. Using a meta-analysis
606 approach, we have shown that while treatment and distribution of water significantly reduces
607 the diversity and abundance of taxa present in the source the subsequent assembly of
608 microbiomes in drinking water is a stochastic process, particularly in the DWDS. This
609 demonstrates that the effects of source and treatment diminish with distance from treatment
610 and time. Only the assembly of microbiomes at the point of chlorination are more
611 deterministic, due to selection pressures on organisms that cannot survive oxidation.
612 Although 16S rRNA amplicon sequencing cannot satisfy current water quality regulation, it
613 can assess the risk from emerging pathogens such as *Legionella* or *Mycobacterium* - which
614 may be in high abundance in DWDS - and track significant changes in the microbiome,
615 which may be associated with contamination or changes in process performance. These are
616 benefits which traditional culture-tests cannot provide. However, further work is required to
617 standardise sequencing and data analysis methods for 16S rRNA amplicon sequencing
618 methods to enable them to be applied within the water industry as standard practice.
619

Acknowledgements

Funding: CT. Scottish Water Industry Funded PhD studentship; CJS Royal Academy of Engineering-Scottish Water Research Chair (RCSRF171864), UZI is supported by NERC, UK, NE/L011956/1.

References

1. Jalava, K. *et al.* Novel microbiological and spatial statistical methods to improve strength of epidemiological evidence in a community-wide waterborne outbreak. *PLoS One* **9**, 104713 (2014).
2. Payment, P. & Locas, A. Pathogens in Water: Value and Limits of Correlation with Microbial Indicators. *Ground Water* **49**, 4–11 (2011).
3. Fricker, C. R. & Eldred, B. J. Identification of coliform genera recovered from water using different technologies. *Lett. Appl. Microbiol.* **49**, 685–688 (2009).
4. Saxena, G., Bharagava, R. N., Kaithwas, G. & Raj, A. Microbial indicators, pathogens and methods for their monitoring in water environment. *J. Water Health* **13**, 319–339 (2015).
5. Leclerc, H., Mossel, D. A. A., Edberg, S. C. & Struijk, C. B. Advances in the Bacteriology of the Coliform Group: Their Suitability as Markers of Microbial Water Safety. *Annu. Rev. Microbiol.* **55**, 201–234 (2001).
6. DWI. Public Drinking Water Quality 2020. *Annual Report* (2020). Available at: https://www.dwi.gov.uk/what-we-do/annual-report/drinking-water-2020/. (Accessed: 17th June 2021)

643 7. Rappé, M. S. & Giovannoni, S. J. The Uncultured Microbial Majority. *Annual Review of*
644 *Microbiology* **57**, 369–394 (2003).
645 8. Hahn, M. W., Koll, U. & Schmidt, J. Isolation and Cultivation of Bacteria. in *The*
646 *Structure and Function of Aquatic Microbial Communities* 313–351 (Springer, Cham,
647 2019). doi:10.1007/978-3-030-16775-2_10
648 9. Savichtcheva, O. & Okabe, S. Alternative indicators of fecal pollution: Relations with
649 pathogens and conventional indicators, current methodologies for direct pathogen
650 monitoring and future application perspectives. *Water Research* **40**, 2463–2476
651 (2006).
652 10. Cabral, J. P. S. Water microbiology. Bacterial pathogens and water. *Int. J. Environ. Res.*
653 *Public Health* **7**, 3657–3703 (2010).
654 11. Lin, W., Yu, Z., Zhang, H. & Thompson, I. P. Diversity and dynamics of microbial
655 communities at each step of treatment plant for potable water generation. *Water*
656 *Res.* **52**, 218–230 (2014).
657 12. Pinto, A. J., Xi, C. & Raskin, L. Bacterial Community Structure in the Drinking Water
658 Microbiome Is Governed by Filtration Processes. *Environ. Sci. Technol.* **46**, 8851–8859
659 (2012).
660 13. Fox, E. J. & Reid-Bayliss, K. S. Accuracy of Next Generation Sequencing Platforms. *J.*
661 *Next Gener. Seq. Appl.* **01**, (2014).
662 14. Vierheilig, J. *et al.* Potential applications of next generation DNA sequencing of 16S
663 rRNA gene amplicons in microbial water quality monitoring. *Water Sci. Technol.* **72**,
664 1962–1972 (2015).
665 15. Han, Z., An, W., Yang, M. & Zhang, Y. Assessing the impact of source water on tap
666 water bacterial communities in 46 drinking water supply systems in China. *Water Res.*

172, 115469 (2020).

16. De Vera, G. A., Gerrity, D., Stoker, M., Frehner, W. & Wert, E. C. Impact of upstream chlorination on filter performance and microbial community structure of GAC and anthracite biofilters. *Environ. Sci. Water Res. Technol.* **4**, 1133–1144 (2018).

17. Gerrity, D. *et al.* Microbial community characterization of ozone-biofiltration systems in drinking water and potable reuse applications. *Water Res.* **135**, 207–219 (2018).

18. Wang, H., Pryor, M. A., Edwards, M. A., Falkinham, J. O. & Pruden, A. Effect of GAC pre-treatment and disinfectant on microbial community structure and opportunistic pathogen occurrence. *Water Res.* **47**, 5760–5772 (2013).

19. Potgieter, S. *et al.* Long-term spatial and temporal microbial community dynamics in a large-scale drinking water distribution system with multiple disinfectant regimes. *Water Res.* **139**, 406–419 (2018).

20. Gülay, A. *et al.* Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. *ISME J.* **10**, 2209–2222 (2016).

21. Shaw, J. L. A. A. *et al.* Using Amplicon Sequencing To Characterize and Monitor Bacterial Diversity in Drinking Water Distribution Systems. *Appl. Environ. Microbiol.* **81**, 6463–6473 (2015).

22. Chan, S. *et al.* Bacterial release from pipe biofilm in a full-scale drinking water distribution system. *npj Biofilms Microbiomes* **5**, 9 (2019).

23. Potgieter, S. C. & Pinto, A. J. Reproducible microbial community dynamics of two drinking water systems treating similar source waters. *bioRxiv Microbiol.* 678920 (2019). doi:10.1101/678920

24. Wu, H. *et al.* Biofilm bacterial communities in urban drinking water distribution systems transporting waters with different purification strategies. *Appl. Microbiol.*

Biotechnol. **99**, 1947–1955 (2014).

25. Prest, E. I. *et al.* Combining flow cytometry and 16S rRNA gene pyrosequencing: A promising approach for drinking water monitoring and characterization. *Water Res.* **63**, 179–189 (2014).

26. Ghaju Shrestha, R. *et al.* Next-generation sequencing identification of pathogenic bacterial genes and their relationship with fecal indicator bacteria in different water sources in the Kathmandu Valley, Nepal. *Sci. Total Environ.* **601**–**602**, 278–284 (2017).

27. McCoy, S. T. & VanBriesen, J. M. Temporal Variability of Bacterial Diversity in a Chlorinated Drinking Water Distribution System. *J. Environ. Eng.* **138**, 786–795 (2012).

28. Bautista-de los Santos, Q. M. *et al.* The impact of sampling, PCR, and sequencing replication on discerning changes in drinking water bacterial community over diurnal time-scales. *Water Res.* **90**, 216–224 (2016).

29. Wolf-Baca, M. & Piekarska, K. Biodiversity of organisms inhabiting the water supply network of Wroclaw. Detection of pathogenic organisms constituting a threat for drinking water recipients. *Sci. Total Environ.* **715**, 136732 (2020).

30. Douterelo, I., Jackson, M., Solomon, C. & Boxall, J. Spatial and temporal analogies in microbial communities in natural drinking water biofilms. *Sci. Total Environ.* **581**–**582**, 277–288 (2017).

31. Bautista-De los Santos, Q. M. *et al.* Emerging investigators series: microbial communities in full-scale drinking water distribution systems – a meta-analysis. *Environ. Sci. Water Res. Technol.* **2**, 631–644 (2016).

32. Zhu, J. *et al.* Mycobacterial metabolic characteristics in a water meter biofilm revealed by metagenomics and metatranscriptomics. *Water Res.* **153**, 315–323 (2019).

715 33. Lührig, K. *et al.* Bacterial community analysis of drinking water biofilms in southern
716 Sweden. *Microbes Environ.* **30**, 99–107 (2015).
717 34. Liu, G. *et al.* Potential impacts of changing supply-water quality on drinking water
718 distribution: A review. *Water Research* **116**, 135–148 (2017).
719 35. Douterelo, I., Sharpe, R. L., Husband, S., Fish, K. E. & Boxall, J. B. Understanding
720 microbial ecology to improve management of drinking water distribution systems.
721 *Wiley Interdiscip. Rev. Water* **6**, e01325 (2019).
722 36. Fish, K. E. & Boxall, J. B. Biofilm Microbiome (Re)Growth Dynamics in Drinking Water
723 Distribution Systems Are Impacted by Chlorine Concentration. *Front. Microbiol.* **9**,
724 2519 (2018).
725 37. Uyaguari-Diaz, M. I. *et al.* Microbial community dynamics of surface water in British
726 Columbia, Canada. *bioRxiv* 719146 (2019). doi:10.1101/719146
727 38. Ahmed, W. *et al.* Toolbox Approaches Using Molecular Markers and 16S rRNA Gene
728 Amplicon Data Sets for Identification of Fecal Pollution in Surface Water. *Appl.*
729 *Environ. Microbiol.* **81**, 7067–7077 (2015).
730 39. Lautenschlager, K. *et al.* Abundance and composition of indigenous bacterial
731 communities in a multi-step biofiltration-based drinking water treatment plant.
732 *Water Res.* **62**, 40–52 (2014).
733 40. Wang, H., Masters, S., Edwards, M. A., Falkinham, J. O. & Pruden, A. Effect of
734 Disinfectant, Water Age, and Pipe Materials on Bacterial and Eukaryotic Community
735 Structure in Drinking Water Biofilm. *Environ. Sci. Technol.* **48**, 1426–1435 (2014).
736 41. Vignola, M., Werner, D., Wade, M. J., Meynet, P. & Davenport, R. J. Medium shapes
737 the microbial community of water filters with implications for effluent quality. *Water*
738 *Res.* **129**, 499–508 (2018).

739 42. Hou, L. *et al.* Spatiotemporal changes in bacterial community and microbial activity in
740 a full-scale drinking water treatment plant. *Sci. Total Environ.* **625**, 449–459 (2018).
741 43. Wan, K. *et al.* Organic carbon: An overlooked factor that determines the antibiotic
742 resistome in drinking water sand filter biofilm. *Environ. Int.* **125**, 117–124 (2019).
743 44. Gensberger, E. T. *et al.* Evaluation of quantitative PCR combined with PMA treatment
744 for molecular assessment of microbial water quality. *Water Res.* **67**, 367–376 (2014).
745 45. Ruike, W., Higashimori, A., Yaguchi, J. & Li, Y. Use of real-time PCR with propidium
746 monoazide for enumeration of viable Escherichia coli in anaerobic digestion. *Water*
747 *Sci. Technol.* **74**, 1243–1254 (2016).
748 46. Liu, J. *et al.* Occurrence and Fate of Ultramicrobacteria in a Full-Scale Drinking Water
749 Treatment Plant. *Front. Microbiol.* **9**, 2922 (2018).
750 47. Liu, G. *et al.* Hotspots for selected metal elements and microbes accumulation and
751 the corresponding water quality deterioration potential in an unchlorinated drinking
752 water distribution system. *Water Res.* **124**, 435–445 (2017).
753 48. Nearing, J. T., Douglas, G. M., Comeau, A. M. & Langille, M. G. I. Denoising the
754 Denoisers: An independent evaluation of microbiome sequence error- correction
755 approaches. *PeerJ* **2018**, (2018).
756 49. Keating, C., Trego, A. C., Sloan, W., O'Flaherty, V. & Ijaz, U. Z. Circular Economy of
757 Anaerobic Biofilm Microbiomes: A Meta-Analysis Framework for Re-exploration of
758 Amplicon Sequencing Data. *bioRxiv* 2020.12.23.424166 (2020).
759 doi:10.1101/2020.12.23.424166
760 50. Leo Lahti, S. S. Introduction to the microbiome R package. (2017). Available at:
761 https://bioconductor.org/packages/devel/bioc/vignettes/microbiome/inst/doc/vigne
762 tte.html. (Accessed: 7th April 2021)

763 51. Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data:
764 Dissimilarity coefficients and partitioning. *Ecol. Lett.* **16**, 951–963 (2013).
765 52. Pinto, A. J., Schroeder, J., Lunn, M., Sloan, W. & Raskin, L. Spatial-Temporal Survey
766 and Occupancy-Abundance Modeling To Predict Bacterial Community Dynamics in
767 the Drinking Water Microbiome. *MBio* **5**, (2014).
768 53. Becerra-Castro, C., Macedo, G., Silva, A. M. T., Manaia, C. M. & Nunes, O. C.
769 Proteobacteria become predominant during regrowth after water disinfection. *Sci.*
770 *Total Environ.* **573**, 313–323 (2016).
771 54. Roeselers, G. *et al.* Microbial biogeography of drinking water: Patterns in
772 phylogenetic diversity across space and time. *Environ. Microbiol.* **17**, 2505–2514
773 (2015).
774 55. Ishii, S., Ksoll, W. B., Hicks, R. E. & Sadowsky, M. J. Presence and growth of naturalized
775 Escherichia coli in temperate soils from lake superior watersheds. *Appl. Environ.*
776 *Microbiol.* **72**, 612–621 (2006).
777 56. Ishii, S. & Sadowsky, M. J. Escherichia coli in the Environment: Implications for Water
778 Quality and Human Health. *microbes and genomes* **23**, 101–108 (2008).
779 57. Ashbolt, N. J. Microbial Contamination of Drinking Water and Human Health from
780 Community Water Systems. *Current environmental health reports* **2**, 95–106 (2015).
781 58. Liu, G. *et al.* Pyrosequencing Reveals Bacterial Communities in Unchlorinated Drinking
782 Water Distribution System: An Integral Study of Bulk Water, Suspended Solids, Loose
783 Deposits, and Pipe Wall Biofilm. *Environ. Sci. Technol.* **48**, 5467–5476 (2014).
784 59. Prest, E. I., Hammes, F., Kötzsch, S., Van Loosdrecht, M. C. M. & Vrouwenvelder, J. S.
785 A systematic approach for the assessment of bacterial growth-controlling factors
786 linked to biological stability of drinking water in distribution systems. *Water Sci.*

Technol. Water Supply **16**, 865–880 (2016).

60. Hassard, F. *Understanding the Use of Flow Cytometry for Monitoring of Drinking Water*. *DWI Reports* (2019).

61. Hammes, F., Goldschmidt, F., Vital, M., Wang, Y. & Egli, T. Measurement and interpretation of microbial adenosine tri-phosphate (ATP) in aquatic environments. *Water Res.* **44**, 3915–3923 (2010).

62. Besmer, M. D. & Hammes, F. Short-term microbial dynamics in a drinking water plant treating groundwater with occasional high microbial loads. *Water Res.* **107**, 11–18 (2016).

63. Props, R., Monsieurs, P., Mysara, M., Clement, L. & Boon, N. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol. Evol.* **7**, 1376–1385 (2016).

64. Favere, J., Buysschaert, B., Boon, N. & De Gusseme, B. Online microbial fingerprinting for quality management of drinking water: Full-scale event detection. *Water Res.* **170**, (2020).