eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Microbial diversity, ecological networks and functional traits associated to materials used in drinking water distribution systems

I. Douterelo[1], Bas E. Dutilh[2,3], K. Arkhipova[2], C. Calero[1], and S. Husband[1]

1. Pennine Water Group, Department of Civil and Structural Engineering, Sir Frederick Mappin Building, University of Sheffield, Sheffield, S1 3JD, UK

2. Theoretical Biology and Bioinformatics, Science for Life, Utrecht University, Hugo R. Kruytgebouw, Padualaan 8, 3584 CH Utrecht, The Netherlands

3. Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Geert Grooteplein Zuid 26-28, 6525 GA, Nijmegen, The Netherlands

Abstract

Drinking water distribution systems host complex microbial communities as biofilms that interact continuously with delivered water. Understanding the diversity, behavioural and functional characteristics will be a requisite for developing future monitoring strategies and protection against water-borne health risks. To improve understanding, this study investigates mobilisation and accumulation behaviour, microbial community structure and functional variations of biofilms developing on different pipe materials from within an operational network. Samples were collected from four pipes during a repeated flushing operation three months after an initial visit that used hydraulic forces to mobilise regenerating biofilms yet without impacting the upstream network. To minimise confounding factors, test sections were chosen with comparable daily hydraulic regimes, physical dimensions, and all connected straight of a common trunk main and within close proximity, hence similar water chemistry, pressure and age. Taxonomical results showed differences in colonising communities between pipe materials, with several genera, including the bacteria *Pseudomonas* and the fungi *Cladosporium,* present in every sample. Diverse bacterial communities dominated compared to more homogeneous fungal, or mycobiome, community distribution. The

analysis of bacterial/fungal networks based on relative abundance of operational taxonomic units (OTUs) indicated microbial communities from cast iron pipes were more stable than communities from the non-ferrous pipe materials. Novel analysis of functional traits between all samples were found to be mainly associated to mobile genetic elements that play roles in determining links between cells, including phages, prophages, transposable elements, and plasmids. The use of functional traits can be considered for development in future surveillance methods, capable of delivering network condition information beyond that of limited conventional faecal indicator tests, that will help protect water quality and public health.

Keywords: biofilms, drinking water, discolouration, fungi-bacteria, functional traits, mobile genetic elements.

Highlights

1. Initial flushing steps yielded samples with highest microbial diversity

2. Pipe material influenced the structure of attached microbial communities

3. Network hydraulics influenced material accumulation and microbial composition

4. The most abundant functional traits were mobile genetic elements

5. The most stable microbial networks were associated with cast iron pipes

1. Introduction

Methodological advances in the characterisation of microorganisms is progressively changing the water industry's view on the microbial ecology of Drinking Water Distribution Systems (DWDS). The attachment of microorganisms that form biofilms on boundary surfaces in DWDS is extensively documented and now considered an explicit part of the system (Douterelo et al. 2016, Liu et al. 2014). Biofilms provide microbial survival advantages over planktonic states, including protection from disinfectants, greater environmental stability and access to nutrient cycling (Donlan 2002). However, biofilms can also act as hotspots for opportunistic pathogens and are linked to water quality deterioration. Thus, understanding biofilm ecology in DWDS and the potential to assess condition will be vital to protect future water quality and ultimately public health (Liu et al. 2017).

Many factors will influence the attachment of microorganisms to drinking water boundary surfaces, including pipe material, hydraulic conditions, nutrient availability and disinfectant residuals (Douterelo et al. 2014a). The hydraulic regime within DWDS is critical in that it drives material exchange between phases, that is bringing material from the bulk water to the pipe surface facilitating accumulation, and when conditions change the subsequent mobilisation back to the bulk water (Abe et al. 2012, Cowle et al. 2014, Douterelo et al. 2013). Processes however are complex, for example,

low flow rates and stagnation may create favourable hydraulic conditions for accumulation of microorganisms and biofilm growth, yet limit nutrient exchange, while varied and higher flow conditions can create adverse hydraulic conditions but with good substrate supply (Douterelo et al. 2013). With biofilm development and the gradual incorporation of low background concentrations of organic and inorganic material, such as lead and other contaminants, gross mobilisation of the biofilm can pose health and water quality risks. A common manifestation of this occurs when flows exceed normal daily values, such as during network interventions or unplanned burst events. In these situations, the increase in shear stress, the force exerted perpendicular to the boundary surface, rapidly mobilises accumulated material resulting in observable discoloured drinking water, higher contaminant concentrations and release of potential opportunistic pathogens (Husband and Boxall 2011).

Although drinking water quality and safety are monitored by water utilities based on indicator microorganisms in bulk water, so a simple pass or fail, surveillance of biofilms or microorganisms during distribution is not part of current regulations. Traditional indicators such as faecal and heterotrophic plate counts are limited regarding their sensitivity, robustness and specificity to detect failures within DWDS. These indicators are also not necessarily correlated with occurrences of new emergent pathogens and are only able to indicate high-intensity disturbances (Saxena et al. 2014). As a consequence, there is a need to develop alternative bioindicators of water quality that can predict, measure and monitor changes within DWDS. Overall this study aims to investigate DWDS physico-chemical characteristics and microbiological genetic features to facilitate evaluation of the potential for such associations to provide viable operational indicators to aid water quality surveillance and management.

The taxonomic characteristics of DWDS microorganisms, mainly bacteria, attached to pipe surfaces has been documented (Douterelo et al. 2014b, Martiny et al. 2003, Revetta et al. 2010). To extend taxonomic knowledge this study includes both analysis of bacteria and fungi in operational DWDS,

since it has been shown that these microorganisms are abundant members of biofilms in DWDS (Douterelo et al. 2018b). There is also sparse information regarding functional traits from these ubiquitous anthropogenic ecosystems, such as identification of the metabolic and physiological features that determine microbial behaviour performance and environmental adaptablity. This limits the ability to understand the relationships established by the metabolically diverse microorganisms within biofilms and environmental factors, such as the impact of pipe surface material on infrastructure performance.

This study analysed microbial communities mobilised from four different pipe materials used by water utilities and includes for the first time characterisation of functional genes. Whole metagenome sequencing was used to characterise potential functional genes in microbiomes with respect to pipe material, hydraulic forces, and associated shifts in water chemistry. Microbial interactions (in the form of correlation networks) in DWDS have also been calculated to indicate the most relevant microbial associations at phylum level. This can provide valuable information about the ecological state (stability) of a given ecosystem and within DWDS can therefore be used to help develop monitoring or maintenance strategies to safeguard the long term delivery of wholesome water.

2. Materials and Methods

2.1 Study sites

Biofilm formation in DWDS can be considered a relatively homogeneous affiliation of microbial and inorganic components defined by the supplied water. Factors such as the pipe surface onto which microorganisms attach and the hydraulics, through the forces it imposes on the 3D structure and driving nutrient transport, will then influence the community composition and functional traits. To investigate these factors, an operational network with different pipe compositions including cast iron (CI), asbestos cement (AC), medium density polyethylene (MDPE) and polyvinyl chloride (uPVC) were subjected to intensive monitoring during two managed step flushing operations conducted three months apart. The first flushing operation in March 2017 tested trial logistics (e.g hydrant

5

accessibility, achievable flow rates, pipe section responses, sample collection, etc.,) and for investigative purposes removed material attached with shear strengths in excess of daily conditions, in essence established biofilm and associated material. The increase in hydraulic forces during flushing, when flows exceed normal conditions, is known to mobilise a mix of organic and inorganic material from pipe walls, as evidenced by changes in turbidity, and this was planned to be collected for analysis (Husband and Boxall. 2011).

The second flushing trial was planned as a copy of the first, with the aim to mobilise material that had regenerated in the three months between visits that would then be collected for detailed microbial analysis. Planned incremental increases in flow facilitated further investigation into changes in material bonding strengths, possibly associated with position or depth of material consolidation on pipe surfaces. The trial network layout and pipe details are shown in Figure 1, highlighting how sections were selected in close proximity and supplied directly from a common trunk main. Ideally water entering each pipe section would be directly comparable, and for this trial the first three sections (MDPE, AC & uPVC) were all supplied directly from a single source treated water and via a 21" AC trunk main. The fourth section, 2 km further downstream with an additional hour transit time, required an additional 18" section of spun iron (SI) trunk main; both trunk mains were laid in 1987. All four trial sections were linked directly to the trunk main, i.e. no intermediary pipe lengths or different pipe materials. This was deemed critical for both operational and study reasons. For the former it allowed increased flows in the trial sections without the risk of producing above normal flows upstream, in this case the supplying trunk main (i.e. no mobilisation impact and therefore potential unmonitored discolouration risk). This also removed the need for any preparatory flushing (and potential network disturbances) and facilitated straightforward return for the second flushing visit. For study reasons it meant the bulk water entering each test section had the same source, treatment and transmission experiences, thereby effectively removing bulk water as an experimental variable. The high demands through the linking trunk main and the close proximity of all test sections as shown in Figure 1 also meant water age and residual chlorine concentrations would be comparable. The test pipes also had

similar diameters, from 75 mm to 125 mm (3" to 5") and were 300 m or more in length, except the MDPE pipe at only 85 m. All pipe sections, with the exception of the uPVC which was a dead-end, supplied downstream networks and experienced daily demand patterns with peak flows of around 10 l/s. Each pipe was over 20 years in service (Figure 1: 85 m of 125 mm MDPE laid 1995; 300 m of 4" (100 mm) AC laid 1962, 420 m of 3" (75 mm) uPVC laid 1967 & 300 m of 4" (100 mm) CI laid 1953) and water company records indicated no flushing or disturbance within the last 10 years, allowing the assumption pipe wall material and communities were in equilibrium with the system and well established. Flushing flows were selected to exceed daily peak, but limit material mobilisation to prevent the particulate turbidity response in the bulk water from producing customer observable discolouration. Prior to the planned mobilising flows being established, a low flow through the standpipe was established and run for 20 minutes or more to allow monitoring equipment to stabilise and ensure all stagnant water from the hydrant was removed. This was confirmed by ensuring residual chlorine was present.

2.2 Online turbidity monitoring and flushing scheme

The flushing involved opening selected hydrants at the end of each pipe section incrementally to achieve target flow rates. To collect flow data during the flushing trials, an ABB Aquamaster flow meter attached to a Langham UK hydrant instrumented standpipe was used. The ABB meter logged flow every 15 seconds with an accuracy of ±5 % of reading and maximum working pressure of 12 bar. Flow (and therefore pipe wall mobilising shear stress forces) was increased via a gate valve connected to the standpipe discharge point. ATI Nephnet instruments with infrared (IR) nephelometric measurement processing were used to measure turbidity responses with dual on-site validation and logging at 1 second. Trials were conducted on the same day between morning and evening demand peaks typical of residential patterns and moving from upstream to downstream sections to retain the same effective 'parcel' of treated water as it travels through the network. A turbidity limit of 4 NTU, below the visual threshold and complying with regulatory requirements at customer taps in the UK, was identified as an upper target for planning purposes that would indicate

7

material mobilisation and provide sufficient concentrations of material for detailed analysis at the sampling hydrant. In normal flushing operations far higher turbidity and discoloured water is commonly experienced, yet with downstream networks beyond the test sections, standard high velocity flushing with the potential to create customer contacts and associated negative consequences was not considered viable (customer contacts are used as a service level indicator in the UK, with target breaches resulting in financial penalties). Each flush was deemed completed when turbidity levels had returned to normal (< 0.5 NTU) and time constrained to ensure all four operations could be completed between morning and evening peak demands.

Water samples at each flushing stage were collected for physico-chemical and microbiological analysis. Several parameters were measured *in-situ* at the times of discrete sample collection prior and during flushing of the pipes. Free and total chlorine were measured with a Palintest method using electrochemical sensors and a Palintest CS 100 Chlorosense Chlorine Meter (Fisher Scientific, UK). Temperature and pH were measured using a Hanna portable meter and probe HI 991003. All other parameters were obtained by accredited laboratory analysis of the discrete samples and using validated methods to meet UK drinking water regulations. Water samples were collected in designated containers and bacterial colony counts were determined by the pour-plate method in which yeast extract agar was used following the UK Environment Agency recommendations (Environment Agency, 2012). Cultures were incubated at 37 °C for 48 h (2-day colony) and 22 °C for 72 h (3-day colony).

2.3 DNA extraction and sequencing

Three replicates of two litres samples of bulk water were taken at the times of discrete sampling during the flushing trials for DNA analysis (total 39 samples). For MDPE 4 flushing steps were set up providing a total of 12 samples and for AC, PVC and CI 3 flushing steps were carried (27 samples), giving a total of 39 samples. These samples were each filtered through 0.22 μm nitrocellulose membrane filters (Millipore, Corp). Filters were preserved in the dark and at -80°C for subsequent

DNA extraction. DNA extraction from filters was carried out by a method based on proteinase K digestion followed by a standard phenol/chloroform-isoamyl alcohol extraction (Douterelo et al., 2013). The quantity and purity of the extracted DNA were assessed using Nanodrop ND-1000 spectrophotometer (NanoDrop, Wilmington, USA).

Sequencing was performed at MR DNA (www.mrdnalab.com, Shallowater, TX, USA) on a MiSeq following the manufacturer's guidelines. The 16S rRNA gene using primers 28F and 519R spanning the V1 to V3 hypervariable regions and primers targeting the ITS1-2 regions were used for bacterial and fungal analysis respectively. These primers were used in a PCR using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions: 94°C for 3 minutes, followed by 30-35 cycles of 94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. After amplification, PCR products were checked in 2 % agarose gel to determine the success of amplification and the relative intensity of bands. Multiple samples were pooled together in equal proportions based on their molecular weight and DNA concentrations. Pooled samples were purified using calibrated Ampure XP beads (Beckman Coulter, Inc.). Then the pooled and purified PCR product were used to prepare the Illumina DNA library. Sequencing data was processed using MR DNA analysis pipeline (MR DNA, Shallowater, TX, USA). In summary, sequences were joined and depleted of barcodes then sequences with less than 150bp and with ambiguous base calls were removed. Sequences were de-noised, and chimeras were detected using UCHIME (Edgar et al. 2011) and removed from further analysis. Final Operational Taxonomic Units (OTUs) were taxonomically classified using BLASTn (Altschul et al. 1990) against a curated database derived from RDPII (Cole et al. 2005) and NCBI. To study alpha-diversity (diversity within samples) the Chao rich estimator, Shannon diversity analysis and Dominance estimators were calculated at 95 % sequence similarity cut off (Douterelo et al., 2013), without rarefaction (McMurdie and Holmes 2014).Selected samples based on peaks of turbidity during flushes, were pooled and subjected to whole metagenome sequencing to have a fingerprint of functional genetic capabilities. The initial concentration of DNA was evaluated using the Qubit® dsDNA HS Assay Kit (Life

Technologies). Because of low DNA concentration whole genome amplification was carried out by using REPLI-g Midi kit (Qiagen). The linear amplified DNA concentration was again evaluated (Table 1) using the Qubit® dsDNA HS Assay Kit (Life Technologies) and 50 ng of DNA was used to prepare the libraries by Nextera DNA Sample preparation (Illumina) following the manufacturer's user guide. The samples underwent the simultaneous fragmentation and addition of adapter sequences. These adapters are utilized during a limited-cycle (5 cycles) PCR in which unique indices were added to the samples. Following the library preparation, the final concentration of the library (Table 1) was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the average library size (Table 1) was determined using the Agilent 2100 Bioanalyzer and reagents (Agilent Technologies, Santa Clara, USA). The libraries were pooled and diluted to 10.0pM and sequenced paired end for 300 cycles using the HiSeq system (Illumina).

The initial library processing involved the individual assessment of quality reads from the 4 libraries using FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality and adapter trimming was performed with BBTools v36.x (https://jgi.doe.gov/data-and-tools/bbtools/) and libraries were assembled individually and cross-assembled with SPAdes v.3.11.0 (Bankevich et al. 2012). Quality of assemblies was assessed with QUAST (Gurevich et al. 2013) and the taxonomical classification of assembled contigs was performed with the CAT (https://github.com/dutilh/CAT) pipeline (von Meijenfeldt et al. 2019). To generate abundance profiles, reads of individual libraries were mapped back to the cross-assembly contigs using bwa mem (Li and Durbin 2009) and the percentage of nucleotides mapped to contigs belonging to each taxon was calculated using SAMtools (Li et al. 2009). Functional annotation of predicted open reading frames on the contigs was extracted from the CAT pipeline results. Additionally, libraries were classified with SEED Subsystems ontological classification system (Overbeek et al. 2005). To connect identifiers of SEED Subsystems to accession identifiers of NCBI nr database, two databases were downloaded on 17 January 2017: nr from NCBI (Benson et al. 2017) and m5nr from MG-RAST (Wilke et al. 2016). A homology search of proteins of the nr database against m5nr was made using

DIAMOND (Buchfink et al. 2014). SEED subsystem annotations were transferred if proteins shared 90 % of amino acid similarity over the full length.

To infer the bacterial and fungal associations in the different pipe materials, microbial ecological networks at phylum level were constructed using SParse InversE Covariance estimation for Ecological Association Inference (SPIEC-EASI) package version 1.0.2. in R (https://github.com/zdk123/SpiecEasi) (Kurtz et al. 2015). For the network construction, the neighbourhood algorithm and the Stability Approach to Regularization Selection (StARS) selection method was performed, using a variability threshold of 0.05. Cytoscape version 3.7.1 (Su et al. 2014) was used to visualize the networks and to perform network-based analysis in order to determine the network connectivity. The analysis provided statistics of the network including density (D) which was calculated using the ratio of the number of edges and the clustering coefficient or transitivity (T) which shows the probability that the close nodes of a node are connected and shows the complexity of a structure (Kim et al. 2018). The ratio between D and T indicates the stability of the network and the lower this ratio, the higher the stability of the network.

2.4 Statistical analyses

To assess the similarity in microbial community structure among samples, the relative sequence abundance at genus level (97 % sequence similarity cut-off) for each sample was used to calculate pairwise similarities. All data were transformed by square root calculations and Bray–Curtis similarity matrixes were generated using the software Primer v6 (PRIMER-E, Plymouth, UK). Bray–Curtis similarity matrixes were visualised using multiple-dimensional scaling (MDS) diagrams. Analysis of similarity statistics (ANOSIM) was calculated using the same Bray–Curtis distance matrix to test the significance of differences among samples based on pipe material for both bacteria and fungi. The values for the ANOSIM R-statistic ranges from -1 to 1, where R = 1 indicates that communities from different pipe materials are completely dissimilar.

To determine if the observed differences in the richness, diversity and dominance indexes were significant, pairwise comparisons between samples from different materials were made using the non-parametric Mann–Whitney U-test using IBM SPSS 21.

3. Results

3.1 Turbidity and metal results

Figure 2 shows the standpipe flushing flows for both the initial and second trials plotted with the turbidity response and iron sample results for each pipe section. Following both trials there were no reports or customer contacts regarding discolouration or other water quality issues. Figure 2 shows how the second visit repeated the initial trial, except for flushing flows increased in the MDPE pipe due to the shorter length in an attempt to mobilise sufficient material for analysis. In trial 1 the turbidity meter used had a range of 0 – 20 NTU, whilst the turbidity meter data from trial 2 was set at 0 – 4 NTU to ensure low range accuracy. All the turbidity results show that for each step increase in flow for both visits there is always further mobilisation of material, indicating that material is attached with a range of shear strengths, and this is the same for re-generating material. On-site operational delays meant to complete the work in time allotted a planned 3-step flushing for each section had to be modified, however this did not detract the experimental aims of observing responses and collecting material samples from each specific pipe in response to elevated mobilising forces.

The MDPE pipe had the shortest temporal turbidity response, as expected since it was the shortest section of pipe flushed. Combining the flushing flow with the background flow (the typical daily demands due to downstream users), typically around 8 l/s during the day in this pipe section, the total flow equates to about 12 l/s or 1 m/s, that is travel times of approximately 85 seconds would be expected and this matches the responses observed (insert Fig. 2). The repeating turbidity response profile combined with travel time analysis indicates that material is being mobilised evenly from over the full pipe length, whilst the increasing magnitude response for equal step increases in flow indicates mobilisation is not a linear correlation to flow rate (Husband and Boxall. 2011). Unusually

in this MDPE section, but seen elsewhere from monitored flushing operations and in the other test sections, there is not a typically observed short-lived initial turbidity spike that accompanies flow increases due to stripping of material from the hydrant fittings.

The first step in the AC pipe flushing in trial 1 produced a noisy response and this may be due to turbidity sample lines becoming partially blocked with small stones from road run-off that had collected in the hydrant fittings (the hydrant was in a natural dip in the road that regularly floods). Small stones also blocked the sample lines during trial 2, but it is unknown if this directly impacted the turbidity results. This is a possibility as results may be considered anomalous when compared to the other 3 sections where turbidity responses during trial 2, representing only 3 months regeneration, are lower than trial 1 after long term undisturbed material accumulation. As for the MDPE pipe, the overall turbidity response during trial 1, excluding the initial noisy spikes and sample line blockage issues, remained lower than the uPVC and CI sections, indicating that both MDPE and AC pipe materials, combined with non-quiescent daily peak flows, may limit material accumulation and pose a lower discolouration risk. Supporting this was metal analysis results and Figure 2 shows iron concentrations significantly lower across all trials in the AC and MDPE sections. In contrast, both the uPVC and CI pipes released greater discolouration material, as shown by the elevated turbidity, and iron sampling results. In the former, the pipe was a dead-end and the rapid increase in turbidity profile, followed by a rapid decrease corresponding to one pipe-length turnover, suggests the high turbidity and iron observed is a result of material sedimentation occurring due to very low daily hydraulics (this pipe supplies only 2 small properties). This behaviour of high quantities of loose material was observed in both steps from the uPVC pipe, suggesting particles with different properties (size, density, composition etc.) had accumulated and were mobilised depending on characteristics in response to the increasing hydraulic forces. Unfortunately, no particle size analysis was undertaken to investigate this. Sample analysis returned high levels of iron from this non-metallic pipe, indicating that although the metal source was likely upstream and transported in the bulk water, as for the MDPE and AC pipes, the accumulation processes where not consistent and this is likely to have an effect on

microbial constitution. The relatively high turbidity during the second trial also suggests the risks associated with sedimentation is greater and returns quicker than processes driving accumulation of material on the pipe walls. The high turbidity observed, in excess of the planned 4 NTU limit, yet produced from similar flow increases to the MDPE and AC sections, indicates zones with low hydraulic demands may be sensitive to minor hydraulic changes.

The CI pipe produced turbidity trends consistent with defined turnover times for each of the flushing stages, indicating material mobilised is from the pipe length under investigation. As expected the second trial returned a lower turbidity response when material had only 3 months to accumulate. With turbidity reaching more than 4 NTU in the third stage during trial 1, the flushing flow (and therefore third stage) was curtailed to prevent further mobilisation, but still maintaining a flow so already mobilised material could be discharged via the standpipe. With the same daily hydraulic patterns as the MDPE and AC pipe sections, the generally higher levels of turbidity observed may be a result of corrosion processes supplementing the bulk water accumulation of material. This is supported by the high levels of iron measured. As noted previously high initial turbidity spikes can be associated with stripping of material in hydrants whenever flow is increased. For the CI monitoring location, a 4 m CI spur with no daily flow was required to connect the hydrant to the distribution pipe. This effective dead-end is responsible for the initial elevated turbidity response at each flow increase, indicating high material accumulation in this short section and likely ongoing corrosion processes with no mechanism to enable material to be removed unless via hydrant operation.

With the planned flow increases and continuous real-time monitoring ensuring controlled material release, no customer complaints for discolouration or other water quality issues were received in relation to any of these flushing operations.

3.2 Discrete sampling and physico-chemical analysis.

 Results from the physico-chemical analysis of the water and the colony counts during the second trail and thereby investigating the re-generation and hence active developing microbial communities

are shown in Table 2. Given the common source supply, most physico-chemical variables were similar among all the samples including; pH (7.7-7.8), conductivity (240 µS/cm), TOC (1.9 mg/l), DOC (2-2.1 mg/l), ammonia (0.003-0.005 mg/l), nitrate (0.52-0.54 mg/l), nitrite (<0.004 mg/l), orthophosphate (<0.03 mg/l), sulphate (8.6-9.5 mg/l) and chlorine (both free and total, 0.6 – 1.1 mg/l). As the shortest section and most upstream, free chlorine (0.8 mg/l) and total chlorine (0.92 mg/l) was slightly higher in the MDPE pipe. Iron concentrations were significantly higher from the CI pipe and also from the PVC pipe samples when compared with other pipe materials.

Colony counts values were only representative in samples from the AC pipe, with up to 7 colonies per ml at 22 °C for 7 days and 3 colonies at 22 °C for 3 days. These however may be a consequence of the material from the hydrant that caused blockages in the sample lines.

3.3 Bacterial and fungal microbial co-occurrence networks

Microbial co-occurrence networks were used to investigate key microbial groups and associations between OTUs at phylum level from the different pipe materials. Figure 3 shows the network analysis for bacteria at phylum level, where each nodule is an OTU, highlighting that each type of pipe material has a different network. Within the network, bacteria belonging to the same phylum tended to present more associations than bacteria from different phyla. Proteobacteria was represented by the majority of OTUs in the network across all types of material, followed by Firmicutes, Actinobacteria and Cyanobacteria. When the stability of the networks was assessed by calculating the ratio between density (D) and clustering coefficient (transitivity, T) of the network was calculated, it was observed that CI (D= 0.018, T=0.117, D/T=0.15) and AC (D= 0.016, T=0.078, D/T=0.20) samples presented the most stable networks and that MDPE (D= 0.010, T=0.034, D/T=0.29) and PVC (D= 0.014, T=0.050, D/T=0.28) samples represented more fragile communities.

For fungi (Fig. 4), the networks for all the materials showed a high number of associations between Ascomycota and Basidiomycota when compared with other phyla. The ratio between the density (D) and clustering coefficient (T) showed that fungal networks were more stable in CI samples (D=0.018,

T=0.120, D/T=0.15) followed by PVC (D=0.021, T=0.125, D/T=0.168), AC (D=0.019, T= 0.110, D/T=0.172) and MDPE samples (D=0.017, T=0.082, D/T=0.207).

3.4 Taxonomic compositional analysis of microbial communities using marker genes: 16S rRNA for bacteria and ITS2 for fungi.

Figure 5 shows the results for the relative abundance of the most abundant OTUs at genus level obtained from the sequencing analysis of marker genes. Independent of the pipe material, the first flushing stage, with the lowest mobilising force, resulted in a richer (higher number of different OTUs) microbial community, both for bacteria and fungi (calculated as the average of the three replicates per sample). Overall, successive flushing stages showed a more homogenous and relatively less rich microbial community when compared with the first flushing step.

The bacterial analysis (Fig. 5A) showed that one of the most represented genera in all samples was *Pseudomonas* with a relative abundance ranging from 4-21 %. Other genera represented in all the samples were *Gloeobacter* and *Staphylococcus*. *Gloeobacter* was more abundant in samples from the first flush (12 %) and decreased with successive flushes (representing an average of 3 %), whilst *Staphylococcus* has an average relative abundance of 3-5 % in all the samples. *Acinetobacter* was the most abundant genus from the MDPE pipe samples, representing 4-5 % of the total microbial community in samples from the initial flushes and increasing above 10 % in the last flushes. The most abundant bacterial genus from material mobilised from the AC pipes was *Staphylococcus* (average 14 %), whilst present on other materials, this genus was particularly abundant in samples from the AC pipes. *Pseudomonas* in AC samples ranged from 4 to 12 % and *Gloeobacter* from 3 to 9 %. In the uPVC pipe, the most abundant genus in all the samples was *Brucella* (1 % first flush to 14 % in the last flush), *Staphylococcus* (from 9 to 5 %) and *Streptococcus* (5 to 12 %). It is important to notice that uPVC was the material where the presence of *Streptococcus* was higher when compared with samples from other materials. From the CI pipe samples, the main characteristics were the occurrence of high relative abundance of *Brucella* (ranging from 6-17 %) and the presence of the genus *Vitis*

(from 10-13 %), which was mainly observed exclusively in these samples. The percentages of these bacterial groups also varied depending on the flushing step.

For fungi (Fig. 5B), samples from the first flushing stage also showed higher richness in all types of pipe material. *Cladosporium* (from 7 to 19 %) was a core genus present in all the samples, being less abundant in CI (from 6 % to 11 %). *Cryptococcus,* was abundant in MDPE (average 12 %, n= 4) and in AC samples (average 16 %, n=3). Other abundant genera were *Aureobasidium*, represented in AC (average 8 %, n=3) and CI (average 3 %, n=3) samples, *Penicillium* in uPVC *(*average 2 %, n=3) and AC (average 4 %, n=3) and *Cadophora* in uPVC *(*average of 1.7 %, n=3) and MDPE (average 4 %, n=4). As with bacteria, compositional shifts in the fungal community structure were also observed between different flushing stages.

Non-metric multidimensional scaling (NMDS) plots of the relative abundance of bacteria show a clear separation of the bacterial community structure at genus level (Fig. 6) between the samples from the different types of pipe material. Note that several outliers were observed due to the natural heterogeneity presented by environmental samples. The analysis of similarities (ANOSIM) confirmed that the bacterial structure was significantly different (Global R= 0.409, p = 0.001). Difference in bacterial structure were particularly significant between AC and uPVC (R= 0.521, p= 0.1 %) and AC and CI (R=0.606, p=0.1 %). Fungal structure (Fig. 6) also showed difference between pipe materials (Global R= 0.206, p=0.1 %), but all the samples tended to cluster together when compared with the bacterial structure. Significant differences in the fungal community structure were only found between MDPE and uPVC (R= 0.317, p= 0.1 %) and MDPE and CI (R= 0.288, p= 0.1 %).

The results for the Shannon diversity index, dominance and Chao1 richness estimator at genus level for bacteria and fungi are shown in Fig. 7. Higher richness and diversity indexes were obtained for bacteria when compared with fungal samples for all the type of pipes. Overall, samples obtained from AC pipes showed the highest diversity (average =4.2, n=9) and CI samples presented least

diversity (average =3.4, n=9), and the difference in richness between these two type of pipes was statistically significant (Mann-Whitney U, p=0.02, 2-tailed). For fungi, higher diversity (Shannon, H) was observed in CI pipe samples (average =2.2, n =9) when compared with samples from other materials. Chao I (richness), was higher for bacteria in AC pipes (average 464.5, n=9) and in uPVC pipes for fungi (average 55.5, n=9). Significant differences in the Shannon index for bacteria were found between MDPE samples and uPVC (Mann-Whitney U, p=0.041, 2-tailed). Dominance was higher for fungi in all samples and material types when compared with bacteria, showing that fewer and more ubiquitous genera were dominating the structural composition of the samples. For fungi, dominance was particularly high for CI samples (average = 0.49, n=9) and within bacteria the highest dominance was presented in uPVC samples (average= 0.31, n=9). Significant differences for bacteria (Mann-Whitney U, p=0.009, 2-tailed) and fungi (Mann-Whitney U, p=0.049, 2-tailed) were found between MDPE and uPVC samples for the dominance indicator.

3.5 Characterisation of genetic functional potential with respect to pipe material.

Table 3 shows the results for the relative abudance of the specific functional annotated categories for genes present in the samples. The major identified SEED categories were virulence, defence and diseases, carbohydrate metabolism and phages, prophages and transposable elements/plasmids. Genes related to virulence, defence and disease had a high percentage in MDPE samples (15 % of the total reads), genes related to carbohydrate metabolism were highly abundant in CI pipes yet presented a very low abundance in AC pipes.

Other abundant genes related to the biological processes of cell wall and capsule functions were highly represented in CI (10.92 %), whilst protein metabolism was high in AC (22.5 %). Genes related with mobile genetic elements (MGE), phages, prophages and transposable elements and plasmids, were highly represented in samples from AC (35 %), CI (24 %), uPVC (12 %) and relatively low in MDPE samples (2 %).

4. Discussion

4.1 Effect of flushing in water physico-chemical characteristics

The turbidity results show that both the uPVC and CI pipe sections in general accumulated more mobilisable material than in the MDPE or AC pipes. Two different mechanisms can be proposed that account for these differences. The first is that hydraulics impact material accumulation, such as when flows are sufficiently low inducing sedimentation of particulate material that otherwise would remain entrained in the bulk flow. At points in the network where this occurs, as in the uPVC dead-end, significant and easily mobilised material can deposit (Husband and Boxall 2011). The second mechanism is the additional material that may be generated by ongoing corrosion processes, as suggested from the CI pipe results (Fig. 2). This is in agreement with previous research carried out in DWDS (Douterelo *et al.*, 2014a; Douterelo *et al.*, 2016), where it was observed that in HDPE and CI pipes there was ubiquitous material accumulation resulting from the background concentrations of material within the bulk water, and that in CI pipes there was additional accumulation of material associated with corrosion.

Flushing, as evidenced by the turbidity response (Fig. 2) to increased flows irrespective of pipe material or hydraulic history, is a valuable strategy to restrict the accumulation of biofilm and of inorganic compounds trapped within them. By careful control of the imposed hydraulics, it is also demonstrated that the rate of material mobilisation can be controlled by operators, limiting the turbidity response and indicating hydraulic control strategies can be effectively applied anywhere within DWDS.

4.2 Influence of flushing and pipe material on the microbiological characteristics of water samples containing material removed from DWDS

This study has characterised the bacterial and fungal communities in material mobilised from an operating DWDS in the UK. Considering the influence of flushing on the material mobilised, the first increase in flow mobilised more material and the developing microbial communities (both bacteria

and fungi) tended to be richer in these samples (Fig. 5 and Fig. 7). Samples from successive flushes showed a more similar microbial structure between them for each type of pipe material. This indicates that patterns in material shear strengths exist where some material is retained and only mobilised with further hydraulic increases. As showed in the NMDS analysis, significant differences in the relative abundance of different microbial groups at genus level were observed between samples from different pipes. The analysis of relative abundance of bacteria showed similarities between the community structure of uPVC and CI samples, and this may be explained by the higher concentrations of iron accumulating in these pipes sections, the former a result of being a dead-end section, the latter ongoing corrosion. From the indexes (Fig. 7), higher bacterial diversity and richness was observed in material detached from AC pipes (calculated average of three replicates) than in samples from other pipe types, but in terms of quantity, higher amounts of material (including particles and biofilm) were detached from CI pipes (Fig. 2). This suggests that in CI pipe samples a less diverse community was present when compared with other pipe materials, with dominant representatives and greater amounts of biofilm and associated particles mobilised. From this it can be suggested that the pipe material influenced the microbial composition, particularly bacteria (Fig. 6) However, it was observed that the accumulation and attachment of material, in terms of quantity, was not associated with higher microbial diversity or richness. Previous research has examined the influence of substrate materials on biofilm development and microbial growth in drinking water related systems (Niquette et al. 2000), showing that bacterial biomass on plastic materials such as PVC and PE were lowest in comparison with iron, with AC intermediate. However, several researchers have shown that plastic, including polyethylene, can stimulate biofilm growth due to the release of biodegradable compounds which can promote microbial growth (Keevil 2003, van der Kooij et al. 2005, Yu et al. 2010). The results from this study indicates that pipe material affects microbial communities, yet other factors such as location in the network and hydraulics, both daily and historic, also determine microorganism surface attachment.

Other factors within the network, more than pipe material, can affect biofilm composition and structure. Regarding the potential effect of pipe age on biofilm communities, it has been previously shown in a 3-year experiment in a model DWDS that mature biofilm (after 256 days of biofilm development), presented a defined and homogeneous bacterial composition (Martiny et al., 2003). The present research studied pipes more than 20 years old and it is expected that mature biofilms are established in this network. Consequently, the microbial diversity of the biofilm attached on the different pipe sections and materials will be comparable in terms of biofilm maturity.

In terms of what type of microorganisms are inhabiting these systems, bacteria showed higher diversity and richness in all pipe materials when compared with fungi (Fig. 7) which had a more homogenous distribution than bacteria. This inter-kingdom interaction, including endosymbiosis or specific cell–cell interactions, can provide microbial communities protection and stability against external stresses and inhibitory compounds (Zhang et al. 2018). The high presence of fungi including species belonging to the phyla Ascomycota and Basidiomycota in DWDS was also observed in drinking water obtained from a desalination plant (Belila et al. 2017). Belila et al. (2017) suggested that materials used to build drinking water infrastructure can leach out biodegradable polymers that fungi growing in pipe sediments and biofilms can degrade. Fungi such as *Penicillium* and *Fusarium*, relatively abundant in this study, are reported to be used in degradation of natural and synthetic polyethylene (Ojha et al. 2017) and have great potential for degradation of different pharmaceutical compounds (Olicon-Hernandez et al. 2017). On the one hand, this potential for biodegradation should be taken into account when selecting the type of material used for building DWDS. On the other hand, fungi and their enzymes capacity for biodegradation could be a potential tool to deal with growing issues such as pharmaceutical compounds, including endocrine disruptors, hormones and antibiotics found in drinking waters systems (Jones et al. 2005). This study confirms that bacterial-fungal consortia are important components of the microbial communities inhabiting all material surfaces in DWDS. Since biofilm formation on pipes is unavoidable, further research is needed to determine if the occurrence of specific inter-kingdom interactions can promote beneficial biofilms to

control and exclude the presence of undesirable microorganisms in DWDS and/or benefit biodegradation of contaminants.

Bacteria were the main drivers of changes in biofilm community structure, in terms of richness and diversity, and this has been observed in other drinking water-related studies (Douterelo et al. 2018b, Revetta et al. 2010, Sun et al. 2014). The most common group of microorganisms in all the samples, but with different relative abundance depending on the pipe material and flushing step, was Proteobacteria. Proteobacteria were represented in this study mainly by the genera *Pseudomonas*, *Acinetobacter* and *Sphingomonas* and in CI pipes *Brucella*. These microorganisms are known inhabitants of drinking water ecosystems (Douterelo et al. 2017, Hou et al. 2018, Simoes et al. 2007) and the network analysis showed that bacteria belonging to Proteobacteria tended to co–occur with each other. The associations between members of the same phylum can be beneficial for the whole biofilm function and contribute to the stability of the community (Boon et al. 2014). Microbial communities can be composed of both stable and variable fractions, the stable fraction has been proposed to be involved in basic functions and the variable fraction to vary rapidly with environmental shifts (Leite et al. 2018). More abundant species in microbial communities tend to form part of the stable fraction and they can influence the presence of other community components (Ma et al. 2016). Consequently, the stability of microbial communities is not entirely determined by external factors and interactions among microorganisms can play a key role in defining biofilm status and resilience (Faust et al. 2015). The co-occurrence patterns are important in understanding microbial community structure and to understand niche spaces shared by community members (Faust and Raes 2012). Resilience refers to the capacity of an ecosystem to recover from a perturbation by the rapid recovering of the microbial community without ending in a different state (Kim et al. 2018). For bacteria, microbial networks obtained from CI samples showed higher stability than for other materials. However, for both bacteria and fungi, MDPE was the least stable network, suggesting that microbial networks on MDPE samples will be least resilient to perturbations (e.g. change in hydraulic

regime, temperature etc.), while CI being more resilient are more likely to form networks similar to those existing prior to perturbations.

From this research it can be suggested that pipe material has a significant influence on microbial community composition. Furthermore, hydraulics within the distribution systems were also affecting the microbiology of the studied system, since the presence of dead-ends in the network influenced the microbial community composition of different pipe material (uPVC and CI). This has been previously overserved by Douterelo et al. 2016, in biofilms of drinking water systems, where very low flow promoted the dominance in the biofilm community of bacterial belonging to the genus Pseudomonas in samples from a ground water supplied DWDS.

4.3 Functional fingerprints and biological processes

Overall, genes related to Mobile Genetic Elements (MGE) were highly abundant in all pipe materials (Table 3). MGE including plasmids, bacteriophages and transposons facilitate DNA transfer between cells (Frost et al. 2005). This transfer normally carries genes that provide their host cells with selective advantages, such as antibiotic resistance, virulence factors or unusual metabolic pathways (Rankin et al. 2011). Within the MGE observed here, plasmids related to function and resistance to antibiotics and toxic compounds were the most abundant annotated genes in biofilms as shown in Table 3 (e.g phages, prophages, transposable elements, and plasmids varied from 2.25 5 in MDPE to 32.32 % in AC).It is known that biofilms in DWDS can house antibiotic resistance bacteria (ARB) independent of the disinfection regime (Douterelo et al. 2018a, Gomez-Alvarez et al. 2016). Zhang *et al.*, 2018, showed using annular reactors that the amount of ARB can increase in tap water due to biofilm detachment. These findings show that it is important to control the mobilisation of material attached to pipes and in particular biofilms to avoid organoleptic problems, but also to protect public health from potential problems associated with antimicrobial resistance.

Gene functional categories documented in this study suggest that the main biological functions associated with biofilm formation are carbohydrate and protein metabolism. In MDPE, genes related to virulence, disease and defence were highly abundant. Some virulence factors can generate new metabolic resources and provide fitness advantages to microorganisms and this can be a strategy to survive in environments with limited oxygen, including biofilms (Hardie 2019). In MDPE samples, genes related to monosaccharides and protein biosynthesis were also highly represented and in AC samples, protein processing and modification and protein degradation and biosynthesis were prevalent (Table 3). However, in CI samples there was less biofilm-related function as much of the material removed was likely to be inorganic including corrosion products and not actual biofilm. In a previous study using samples from an HDPE pipe in an operating DWDS (Douterelo et al., 2018), whole metagenomics revealed habitat-specific (biofilm vs. water) differences in terms of gene functions. Gene functional categories documented by Douterelo et al., 2018 analysed with MGRAST and annotated using COGs suggested that the main metabolic pathways associated with biofilm formation are polysaccharide biosynthesis and sialic acid synthesis. Furthermore, several resistance mechanisms were identified preferentially within biofilms, including genes associated with the prevention and repair of disinfectant radical-induced damage and antibiotic resistance. These findings corroborate our previous understanding that rather than intervening at the water treatment works to control water quality, it might be necessary to invest in network interventions, such as cleaning strategies or probiotic approaches to prevent adverse biofilm characteristics and ultimately protect public health.

As illustrated by these results, our metagenomic sampling approach was able to identify diverse microbial community members from different kingdoms and potential functional traits by bulk isolation of all DNA in the system. As a consequence, at this point we did not distinguish extracellular DNA (eDNA) from DNA that was present in living microbial cells. While eDNA may form a significant component of the nucleic acids in a given environment (Sakcham et al 2019), we expect that the most abundant community members will also contribute to the majority of eDNA. Thus,

while the effect of eDNA may be limited, it will be interesting to distinguish these compartments by follow-up experiments in the future.

This innovative research, investigating the impact of pipe material and hydraulic significance on developing biofilms, highlights comparisons and differences in microbial diversity, ecological networks and functional traits. With current faecal indicator tests only providing risk detection, this work improves understanding and highlights the potential for new approaches to inform future monitoring or control strategies to protect drinking water quality.

5. Conclusions

This study applied DNA sequencing (bacteria and fungi) and whole metagenomics sequencing to samples mobilised from different pipe materials within the same network and fed by the same source water during planned flushing trials with the aim to characterise microbial communities and compare environmental responses, taking into account taxonomic and functional traits. The results indicate:

- Pipe material influences microbial community, functional traits and stability.

- Bacterial/fungal network analysis, based on OTUs relative abundance and statistical correlation, indicate the most stable communities were associated with cast iron pipe samples.

- Metagenomics can be used to better understand the microbial ecology of DWDS and aid the monitoring of processes in these engineered systems.

- Improved understanding with monitoring and analysis can promote the implementation of ecological approaches towards the management of the DWDS microbiome (such as controlled hydraulic management to remove weakly attached communities) to prevent system failures and protect public health.

REFERENCES

Abe, Y., Skali-Lami, S., Block, J.-C. and Francius, G. (2012) Cohesiveness and hydrodynamic properties of young drinking water biofilms. Water Research 46(4), 1155-1166.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J Mol Biol 215(3), 403-410.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5), 455-477.

Belila, A., El-Chakhtoura, J., Saikaly, P.E., van Loosdrecht, M.C.M. and Vrouwenvelder, J.S. (2017) Eukaryotic community diversity and spatial variation during drinking water production (by seawater desalination) and distribution in a full-scale network. Environmental Science: Water Research & Technology 3(1), 92-105.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K.D. and Sayers, E.W. (2017) GenBank. Nucleic Acids Research 46(D1), D41-D47.

Boon, E., Meehan, C.J., Whidden, C., Wong, D.H.-J., Langille, M.G.I. and Beiko, R.G. (2014) Interactions in the microbiome: communities of organisms and communities of genes. FEMS Microbiology Reviews 38(1), 90-118.

Buchfink, B., Xie, C. and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. Nature methods 12, 59.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M. and Tiedje, J.M. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Research 33(Database issue), D294-D296.

Cowle, M.W., Babatunde, A.O., Rauen, W.B., Bockelmann-Evans, B.N. and Barton, A.F. (2014) Biofilm development in water distribution and drainage systems: dynamics and implications for hydraulic efficiency. Environmental Technology Reviews 3(1), 31-47.

Donlan, R.M. (2002) Biofilms: Microbial life on surfaces. Emerging Infectious Diseases 8(9), 881-890.

Douterelo, I., Boxall, J.B., Deines, P., Sekar, R., Fish, K.E. and Biggs, C.A. (2014a) Methodological approaches for studying the microbial ecology of drinking water distribution systems. Water Research 65, 134-156.

Douterelo, I., Calero-Preciado, C., Soria-Carrasco, V. and Boxall, J.B. (2018a) Whole metagenome sequencing of chlorinated drinking water distribution systems. Environmental Science: Water Research & Technology.

Douterelo, I., Fish, K.E. and Boxall, J.B. (2018b) Succession of bacterial and fungal communities within biofilms of a chlorinated drinking water distribution system. Water Research 141, 74-85.

Douterelo, I., Jackson, M., Solomon, C. and Boxall, J. (2016) Microbial analysis of in situ biofilm formation in drinking water distribution systems: implications for monitoring and control of drinking water quality. Appl Microbiol Biotechnol 100(7), 3301-3311.

Douterelo, I., Jackson, M., Solomon, C. and Boxall, J. (2017) Spatial and temporal analogies in microbial communities in natural drinking water biofilms. Science of the Total Environment 581, 277-288.

Douterelo, I., Sharpe, R. and Boxall, J. (2014b) Bacterial community dynamics during the early stages of biofilm formation in a chlorinated experimental drinking water distribution system: implications for drinking water discolouration. Journal of Applied Microbiology.

Douterelo, I., Sharpe, R.L. and Boxall, J.B. (2013) Influence of hydraulic regimes on bacterial community structure and composition in an experimental drinking water distribution system. Water Research 47(2), 503-516.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16), 2194-2200.

Faust, K., Lima-Mendez, G., Lerat, J.-S., Sathirapongsasuti, J.F., Knight, R., Huttenhower, C., Lenaerts, T. and Raes, J. (2015) Cross-biome comparison of microbial association networks. Frontiers in Microbiology 6, 1200.

Faust, K. and Raes, J. (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10(8), 538-550.

Frost, L.S., Leplae, R., Summers, A.O. and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3(9), 722-732.

Gomez-Alvarez, V., Pfaller, S., Pressman, J.G., Wahman, D.G. and Revetta, R.P. (2016) Resilience of microbial communities in a simulated drinking water distribution system subjected to disturbances: role of conditionally rare taxa and potential implications for antibiotic-resistant bacteria. Environmental Science: Water Research & Technology 2(4), 645-657.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8), 1072-1075.

Hardie, K.R. (2019) Where bacterial metabolism and virulence intersect. Microbiology Today 46(2), 58-61.

Hou, L., Zhou, Q., Wu, Q., Gu, Q., Sun, M. and Zhang, J. (2018) Spatiotemporal changes in bacterial community and microbial activity in a full-scale drinking water treatment plant. Science of the Total Environment 625, 449-459.

Husband, P.S. and Boxall, J.B. (2011) Asset deterioration and discolouration in water distribution systems. Water Research 45(1), 113-124.

Jones, O.A., Lester, J.N. and Voulvoulis, N. (2005) Pharmaceuticals: a threat to drinking water? Trends in Biotechnology 23(4), 163-167.

Keevil, C.W. (2003) Rapid detection of biofilms and adherent pathogens using scanning confocal laser microscopy and episcopic differential interference contrast microscopy. Water Science and Technology 47(5), 105-116.

Kim, H.-J., Kim, H., Kim, J.J., Myeong, N.R., Kim, T., Park, T., Kim, E., Choi, J.-y., Lee, J., An, S. and Sul, W.J. (2018) Fragile skin microbiomes in megacities are assembled by a predominantly niche-based process. Science Advances 4(3), e1701581.

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J. and Bonneau, R.A. (2015) Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Computational Biology 11(5), e1004226.

Leite, D.C.A., Salles, J.F., Calderon, E.N., van Elsas, J.D. and Peixoto, R.S. (2018) Specific plasmid patterns and high rates of bacterial co-occurrence within the coral holobiont. Ecology and Evolution 8(3), 1818-1832.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16), 2078-2079.

Liu, G., Tao, Y., Zhang, Y., Lut, M., Knibbe, W.-J., van der Wielen, P., Liu, W., Medema, G. and van der Meer, W. (2017) Hotspots for selected metal elements and microbes accumulation and the corresponding water quality deterioration potential in an unchlorinated drinking water distribution system. Water Research 124, 435-445.

Liu, R., Zhu, J., Yu, Z., Joshi, D., Zhang, H., Lin, W. and Yang, M. (2014) Molecular analysis of long-term biofilm formation on PVC and cast iron surfaces in drinking water distribution system. Journal of Environmental Sciences 26(4), 865-874.

Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., Brookes, P.C., Xu, J. and Gilbert, J.A. (2016) Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. The Isme Journal 10, 1891.

Martiny, A.C., Jørgensen, T.M., Albrechtsen, H.-J., Arvin, E. and Molin, S. (2003) Long-Term Succession of Structure and Diversity of a Biofilm Formed in a Model Drinking Water Distribution System. Applied and Environmental Microbiology 69(11), 6899-6907.

McMurdie, P.J. and Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLOS Computational Biology 10(4), e1003531.

Niquette, P., Servais, P. and Savoir, R. (2000) Impacts of pipe materials on densities of fixed bacterial biomass in a drinking water distribution system. Water Research 34(6), 1952-1956.

Ojha, N., Pradhan, N., Singh, S., Barla, A., Shrivastava, A., Khatua, P., Rai, V. and Bose, S. (2017) Evaluation of HDPE and LDPE degradation by fungus, implemented by statistical optimization. Scientific Reports 7, 39515.

Olicon-Hernandez, D.R., Gonzalez-Lopez, J. and Aranda, E. (2017) Overview on the Biochemical Potential of Filamentous Fungi to Degrade Pharmaceutical Compounds. Front Microbiol 8(1792).

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33(17), 5691-5702.

Rankin, D.J., Rocha, E.P. and Brown, S.P. (2011) What traits are carried on mobile genetic elements, and why? Heredity 106(1), 1-10.

Revetta, R.P., Pemberton, A., Lamendella, R., Iker, B. and Santo Domingo, J.W. (2010) Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. Water Research 44(5), 1353-1360.

Sakcham, B., et al. (2019). Extracellular DNA in Monochloraminated Drinking Water and Its Influence on DNA-Based Profiling of a Microbial Community. Environmental Science & Technology Letters 6(5): 306-312.

Saxena, G., Bharagava, R.N., Kaithwas, G. and Raj, A. (2014) Microbial indicators, pathogens and methods for their monitoring in water environment. Journal of Water and Health 13(2), 319-339.

Simoes, L.C., Simoes, M. and Vieira, M.J. (2007) Biofilm interactions between distinct bacterial genera isolated from drinking water. Appl Environ Microbiol 73(19), 6192-6200.

Su, G., Morris, J.H., Demchak, B. and Bader, G.D. (2014) Biological network exploration with Cytoscape 3. Current protocols in bioinformatics 47, 8.13.11-18.13.24.

Sun, H., Shi, B., Bai, Y. and Wang, D. (2014) Bacterial community of biofilms developed under different water supply conditions in a distribution system. Science of the Total Environment 472, 99-107.

van der Kooij, D., Veenendaal, H.R. and Scheffer, W.J.H. (2005) Biofilm formation and multiplication of Legionella in a model warm water system with pipes of copper, stainless steel and cross-linked polyethylene. Water Research 39(13), 2789-2798.

von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H. and Dutilh, B.E. (2019) Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. bioRxiv, 530188.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P., Paczian, T., Trimble, W.L., Bagchi, S., Grama, A., Chaterji, S. and Meyer, F. (2016) The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Research 44(D1), D590-D594.

Yu, J., Kim, D. and Lee, T. (2010) Microbial diversity in biofilms on water distribution pipes of different materials. Water Science and Technology 61(1), 163-171.

Zhang, J., Li, W., Chen, J., Qi, W., Wang, F. and Zhou, Y. (2018) Impact of biofilm formation and detachment on the transmission of bacterial antibiotic resistance in drinking water distribution systems. Chemosphere 203, 368-380.

Figure 1: Trial sites identification in Wales (UK), location and pipe properties in the monitored chlorinated network.
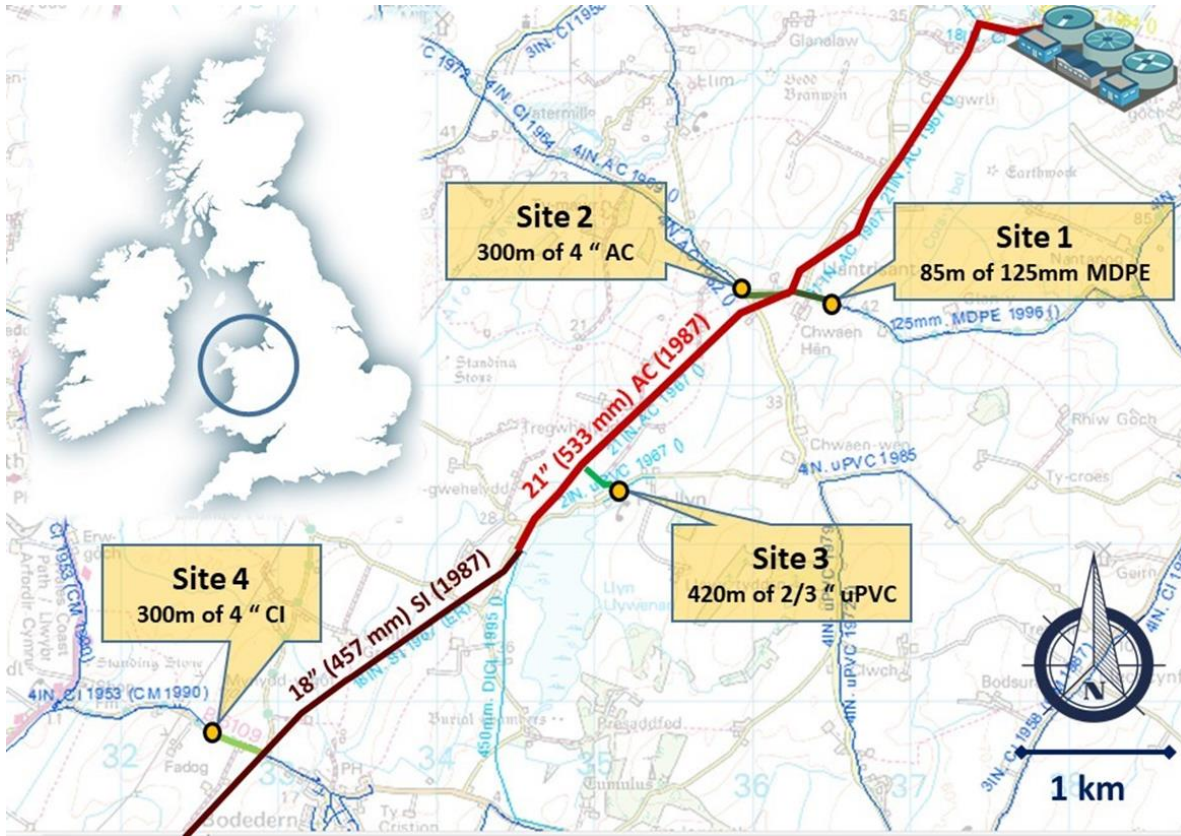
Figure 2: Trial sites identification, location and pipe properties.
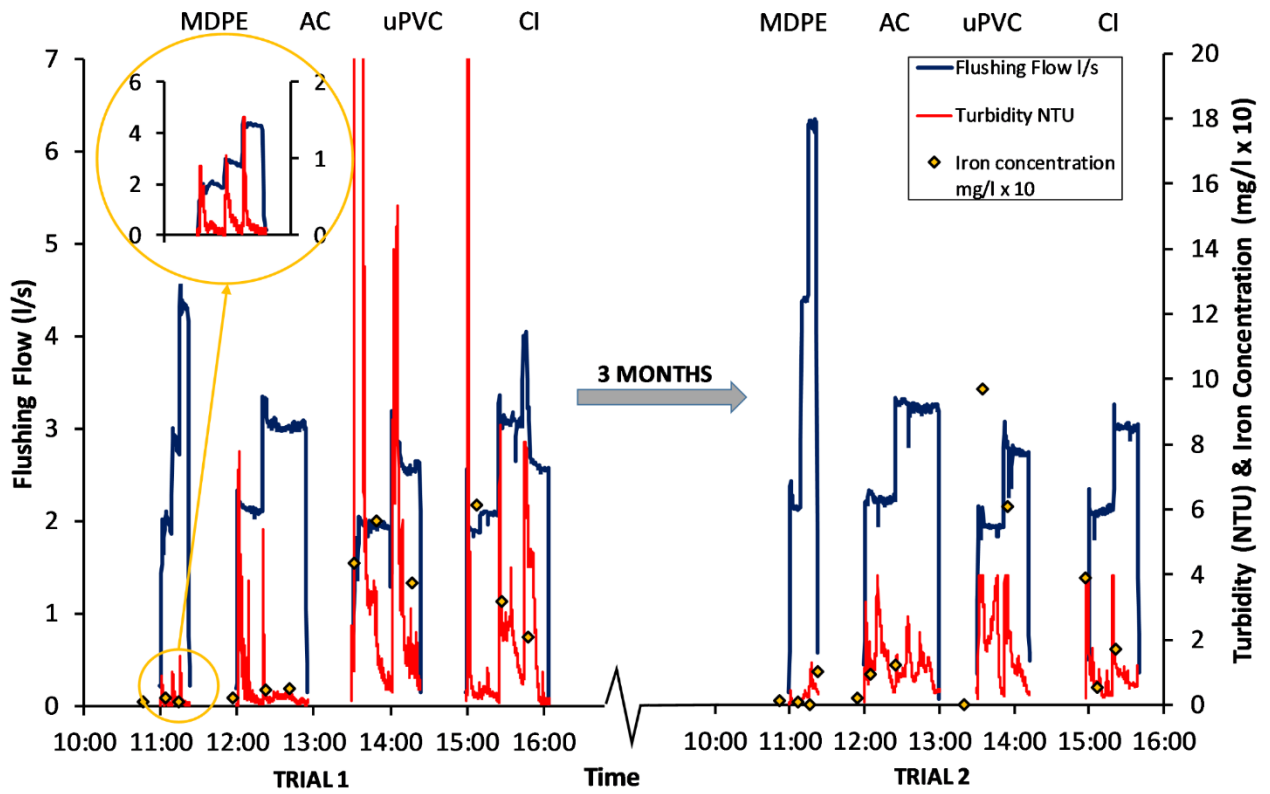
Figure 3: Bacterial networks analysis at phylum level constructed using SParse InversE Covariance estimation for different type of pipe materials; PVC, Asbestos Cement (AC), Medium-density polyethylene (MDPE) and Cast iron (CI) (n=39). The statistical analysis of the networks includes results showing T = transitivity (clustering coefficient) and D= density, calculated using the ratio of the number of edges. Each node represents OTUs at phylum level.
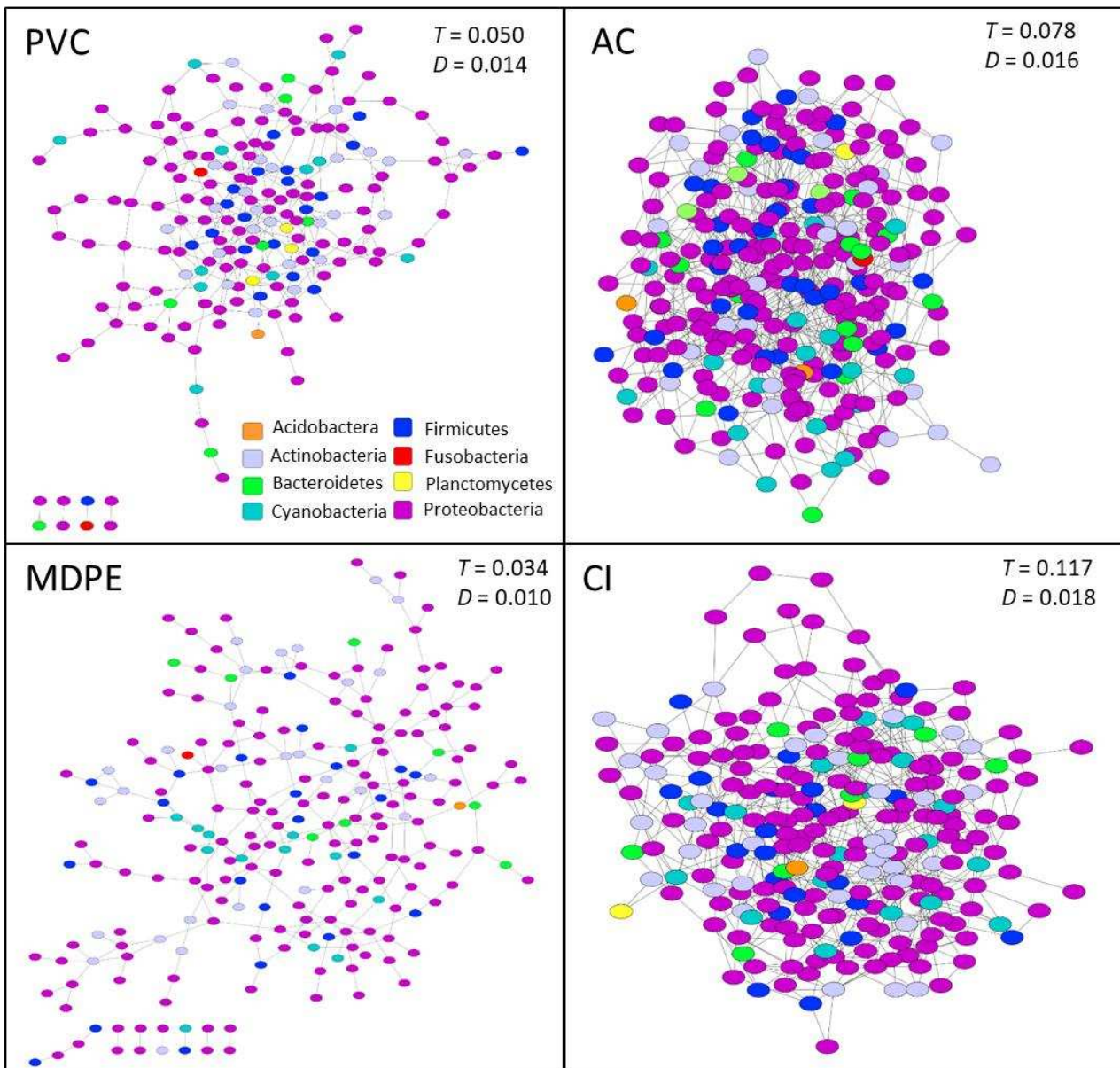
Figure 4: Fungal networks analysis at phylum level constructed using SParse InversE Covariance estimation for different type of pipe materials PVC, Asbestos Cement (AC), Medium-density polyethylene (MDPE) and Cast iron (CI) (n=39). The statistical analysis of the networks includes results showing T = transitivity (clustering coefficient) and D= density, calculated using the ratio of the number of edges. Each node represents OTUs at phylum level.
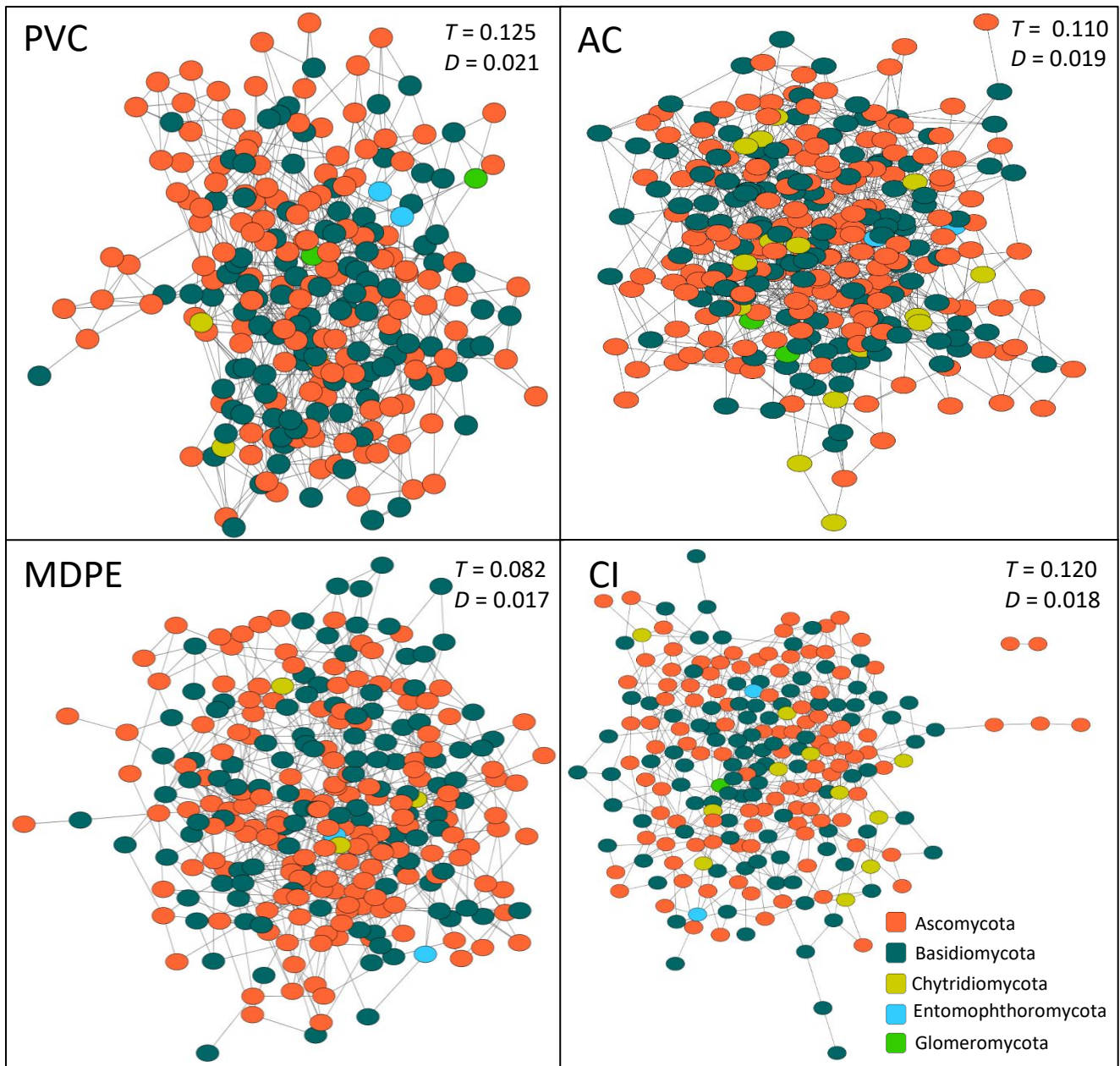
Figure 5: Percentage of relative abundance of A) bacteria and B) fungi at genus level for each of the flushing steps (average of 3 replicates, n=3) and materials. Numbers on the pipe materials labels indicate the flushing step. The first flushing step is indicated with 1*.
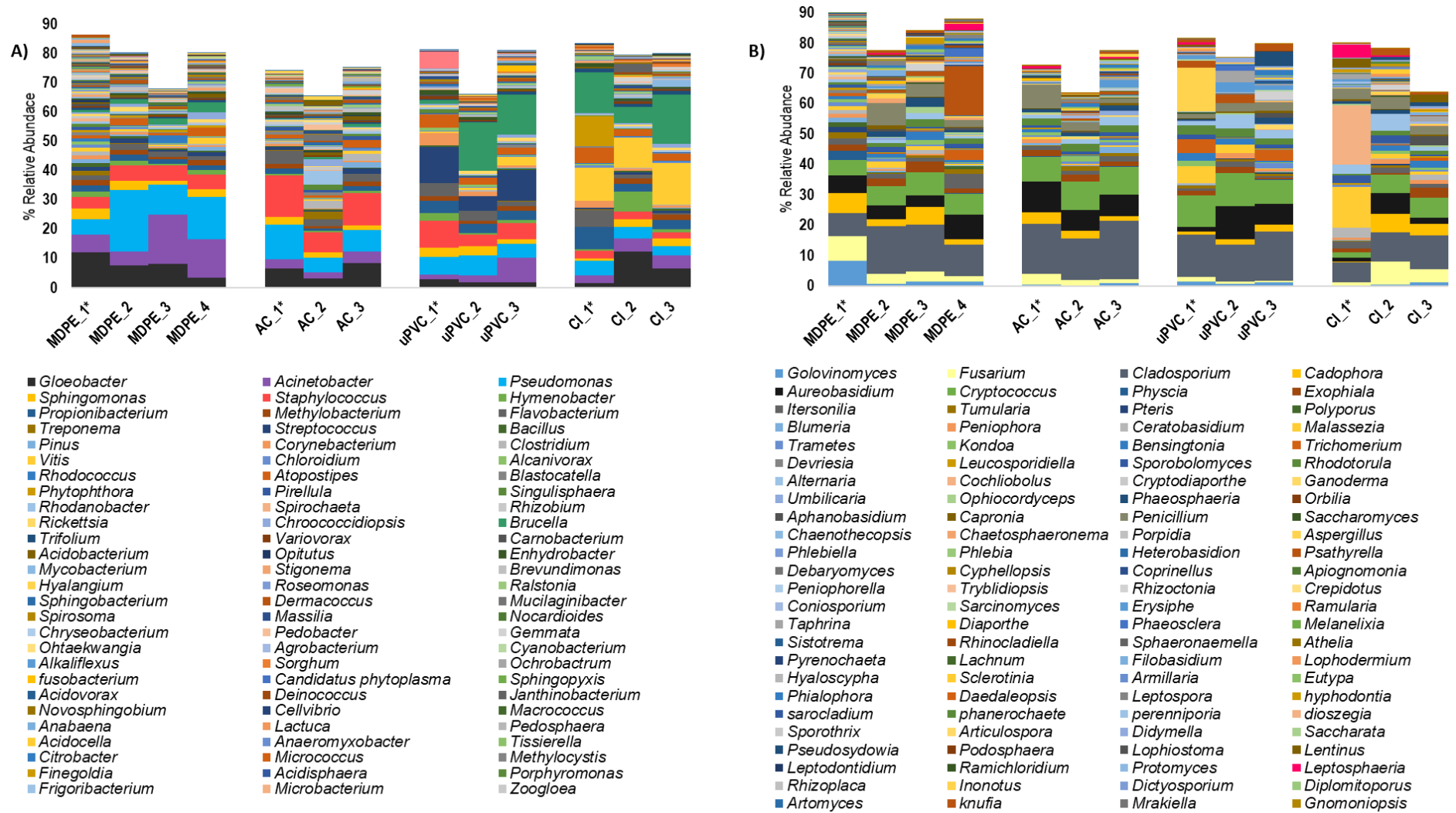
Figure 6: Two-dimensional plot of the multidimensional scaling (MDS) analysis based on Bray–Curtis similarities of the percentage sequence abundance (A) bacteria and (B) fungi. Showing differences in the bacterial community structure between polyethylene and cast iron biofilms (n = 39). Symbols are representing individual samples and are coloured based on sample type.
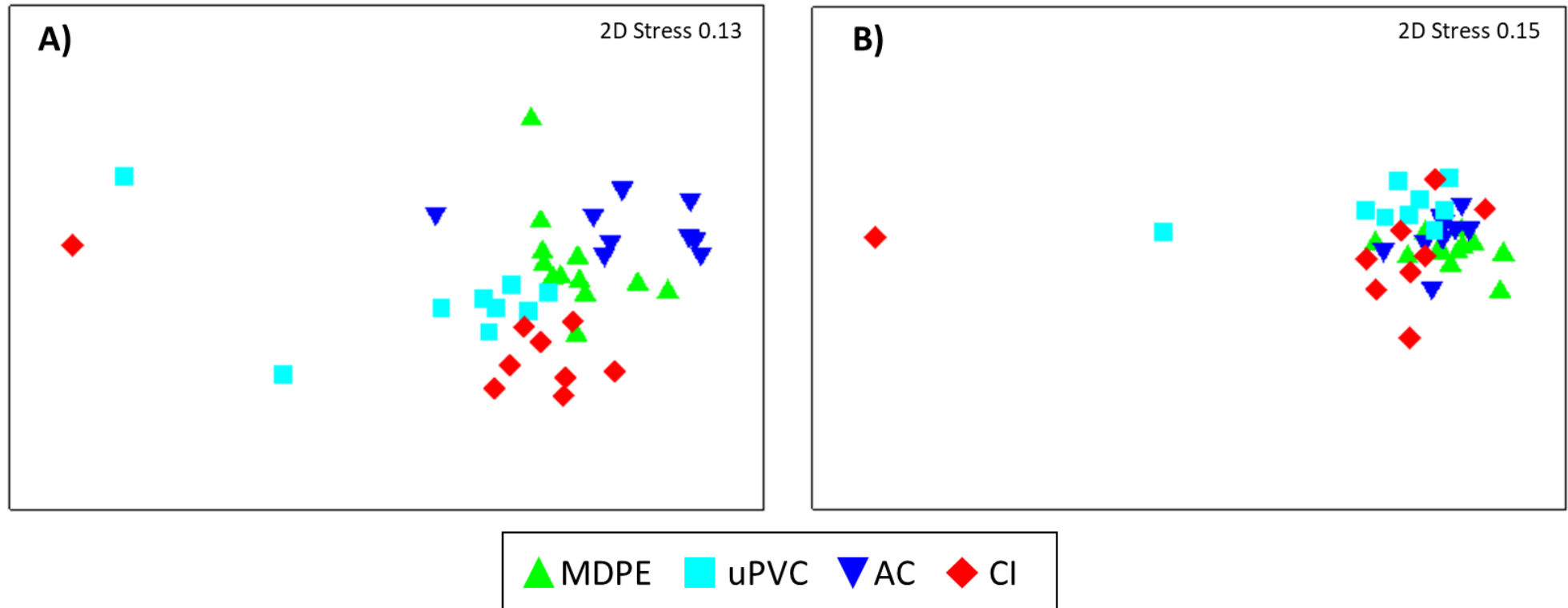
Figure 7 : Diversity index (Shannon), Chao I (richnes index) and Dominance for OTUs at 95 % cut off for bacteria (blue markers) and fungi (orange markers). Names on axis Y are pipe material (MDPE, AC, PVC and CI), numbers are flushing step (1 to 4) and number of replicate (1,2,3).
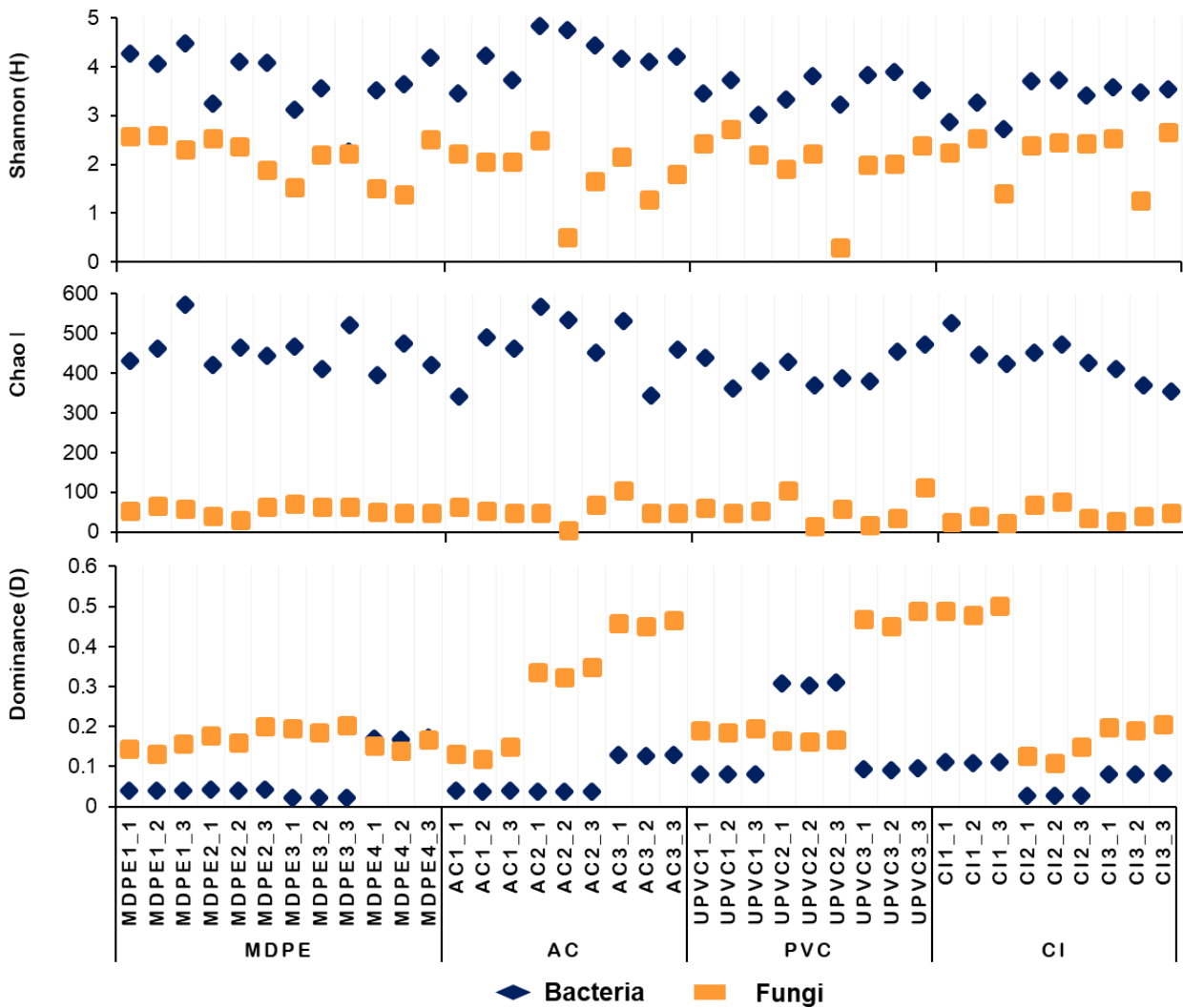
Table 1. Initial concentration of DNA in samples, linear amplified DNA, final library concentration, and average library size.

Site ID & pipe material	DNA concentration (ng/µL)	*DNA concentration (ng/µL)	Final library DNA concentration (ng/µL)	Average library size (bp)
1: MDPE	42.70	298.00	2.74	967
2: AC	190.30	380.00	2.38	1097
3: uPVC	83.20	504.00	4.62	1184
4: CI	80.16	628.00	11.90	1346
	Linear amplified DNA			

Table 2: Water physico-chemical analysis and microbial-culture dependent results of discrete water samples

Site	Material	pH	Conductivity 20 °C (µS/cm)	NTU	Mn (mg/l)	Fe(mg/l)	TOC (mg/l)	Ammonia (mg/l)	Nitrate (mg/l)	Nitrite (mg/l)	Ortophosphate as P (mg/l)	Sulphate (mg/l)	Lead (mg/l)	Chlorine (free) (mg/l)	Chlorine (total) (mg/l)	Colonies 7 days 22 °C (c/ml)	Colonies 3 Days 22 °C (c/ml)	Colonies 2 Days 37° C (c/ml)
1	MDPE	7.7	240	0.11	0.012	0.014	1.9	0.003	0.54	<0.004	<0.03	9.1	0.00018	0.76	0.76	2	0	0
1	MDPE	7.7	240	0.06	0.012	0.007	1.9	0.005	0.54	<0.004	<0.03	9.2	0.00009	0.76	0.98	0	0	0
1	MDPE	7.7	240	0.18	0.016	*	2.0	0.005	0.55	<0.004	<0.03	9.2	0.00013	0.72	0.74	0	0	0
1	MDPE	7.7	240	0.66	0.035	0.100	1.9	0.004	0.52	<0.004	<0.03	9.1	0.00019	1.01	1.10	0	2	0
2	AC	7.8	240	0.06	0.013	0.020	1.9	0.004	0.55	<0.004	<0.03	9.5	0.00007	0.40	0.60	7	4	0
2	AC	7.8	240	0.37	0.014	0.093	1.9	0.003	0.52	<0.004	<0.03	9	0.0003	0.60	1.10	1	0	1
2	AC	7.7	240	0.54	0.018	0.120	1.9	0.003	0.53	<0.004	<0.03	9.3	0.00011	0.85	1.00	0	0	0
3	PVC	7.7	240	0.20	0.013	*	1.9	0.003	0.54	<0.004	<0.03	9.1	0.00005	0.73	0.90	0	0	0
3	PVC	7.7	240	4.40	0.053	0.970	1.9	0.003	0.52	<0.004	<0.03	9.1	0.0014	0.80	0.80	0	1	0
3	PVC	7.7	240	1.80	0.029	0.610	1.9	0.003	0.53	<0.004	<0.03	9.1	0.00066	0.87	0.88	0	0	0
4	CI	7.7	240	3.00	0.013	0.390	1.9	0.003	0.53	<0.004	<0.03	9.2	0.00005	0.68	0.92	3	0	0
4	CI	7.7	240	0.29	0.018	0.054	1.9	0.003	0.52	<0.004	<0.03	9.8	0.00005	0.93	0.92	0	0	1
4	CI	7.7	240	0.70	0.019	0.170	1.9	0.003	0.53	<0.004	<0.03	8.6	0.00005	0.75	1.00	0	0	0

* No data obtained

Table 3: Percentage of abundance of the most relevant genetic functions from the different pipe materials based on SEED ontological classification.
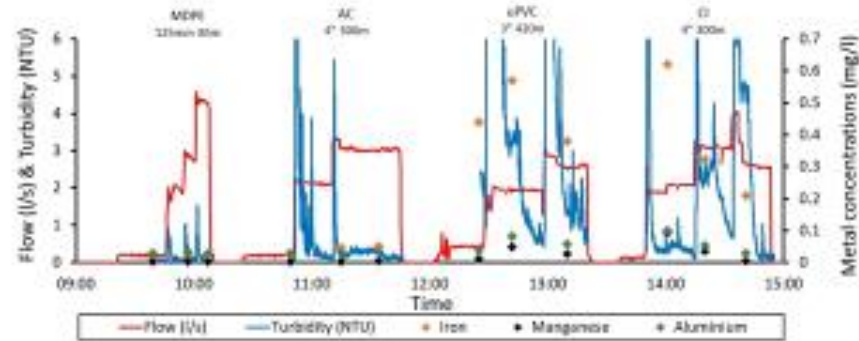
Biological	MDP	A	uPV	C
Virulence, Disease and	15.3	3.4	11.4	5.8
Carbohydrat	12.6	0.8	5.3	23.9
Clustering-based	12.6	2.0	13.8	5.8
Regulation and Cell	9.6	3.4	0.6	0.3
Potassium	8.5	5.2	0.7	0.3
Cell Wall and	6.8	1.1	6.6	10.9
Amino Acids and	5.5	1.6	1.1	1.2
Protein	5.0	22.0	6.5	1.0
RNA	4.2	1.0	8.6	1.7
Membrane	4.0	3.6	5.0	2.1
Phosphorus	2.8	0.8	0.0	0.9
Stress	2.7	0.1	2.2	5.8
Nucleosides and	2.5	0.4	0.8	0.1
Phages, Prophages, Transposable elements,	2.2	35.3	12.1	23.7
Fatty Acids, Lipids, and	2.1	0.4	9.9	0.2
Photosynthe	1.3	N	0.5	0.1
DNA	0.5	0.5	4.8	0.6
Respirati	0.4	15.8	0.6	10.0
Cell Division and Cell	0.3	N	0.0	0.0
Cofactors, Vitamins, Prosthetic Groups,	0.1	0.0	2.6	0.7
Metabolism of Aromatic	0.1	0.0	0.1	1.0
Miscellaneo	0.0	1.7	0.0	0.0
Motility and	0.0	N	0.0	0.5
Iron acquisition and	0.0	N	2.7	2.2
Nitrogen	0.0	0.0	0.2	0.1
Secondary	0.0	N	0.0	0.0
Sulfur	0.0	N	2.7	0.0

NA= no