# Medium shapes the microbial community of water filters with implications for effluent quality

Marta Vignola [a, c, *], David Werner [a], Matthew J. Wade [a], Paola Meynet [a, b], Russell J. Davenport [a]

[a] School of Engineering, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom
[b] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstr. 133, Dübendorf, CH-8600, Switzerland
[c] College of Science and Engineering, Division of Infrastructure and Environment, School of Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom

## ARTICLE INFO

## ABSTRACT

Little is known about the forces that determine the assembly of diverse bacterial communities inhabiting drinking water treatment filters and how this affects drinking water quality. Two contrasting ecological theories can help to understand how natural microbial communities assemble; niche theory and neutral theory, where environmental deterministic factors or stochastic factors predominate respectively. This study investigates the development of the microbial community on two common contrasting filter materials (quartz sand and granular activated carbon-GAC), to elucidate the main factors governing their assembly, through the evaluation of environmental (i.e. filter medium type) and stochastic forces (random deaths, births and immigration). Laboratory-scale filter columns were used to mimic a rapid gravity filter; the microbiome of the filter materials, and of the filter influent and effluent, was characterised using next generation 16S rRNA gene amplicon sequencing and flow-cytometry. Chemical parameters (i.e. dissolved organic carbon, trihalomethanes formation) were also monitored to assess the final effluent quality. The filter communities seemed to be strongly assembled by selection rather than neutral processes, with only 28% of those OTUs shared with the source water detected on the filter medium following predictions using a neutral community model. GAC hosted a phylogenetically more diverse community than sand. The two filter media communities seeded the effluent water, triggering differences in both water quality and community composition of the effluents. Overall, GAC proved to be better than sand in controlling microbial growth, by promoting higher bacterial decay rates and hosting less bacterial cells, and showed better performance for putative pathogen control by leaking less *Legionella* cells into the effluent water.

## 1. Introduction

Bacterial communities are present with high abundance in drinking water treatment systems, not only in source waters ($10^5$-$10^6$ cells/ml) (Hammes et al., 2008; Lautenschlager et al., 2014) and at different points of a Drinking Water Treatment Plant (DWTP) ($10^5$ cells/ml) (Hammes et al., 2008; Vital et al., 2012) but also in the distribution networks ($10^4$-$10^5$ cells/ml) (Hammes et al., 2008; Lautenschlager et al., 2013). Modern DWTPs employ multi-step treatment processes to bring source water up to acceptable drinking water standards. Among them, filters are widely used due to their high performance in removing fine solids and soluble organic matter, their low energy consumption and low maintenance requirements.

Water filters are populated by a high diversity of bacteria (Pinto et al., 2012), which may contribute in the removal of contaminants, such as dissolved organic carbon (DOC), nitrogen, and micropollutants (Richter et al., 2008), or might play a detrimental role, by harbouring potential pathogens or releasing DOC, with a subsequent increase in trihalomethane (THM) formation. Moreover, water filters have a significant influence on the bacterial community composition in the post-filtration water and distribution network (Pinto et al., 2012).

* Corresponding author. School of Engineering, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom.
E-mail address: marta.vignola@glasgow.ac.uk (M. Vignola).

Nevertheless, little is known about how water treatment bacterial communities assemble, what forces determine their structure in water filters, and how this affects their functions, such as the quality of the final drinking water, in water treatment processes. Better identification of these forces provides a challenging opportunity to optimise DWTPs performance.

Two contrasting ecological theories explain how natural microbial communities assemble: the traditional niche-based theory (Ramette and Tiedje, 2007; Tilman, 1982) and the unified neutral theory of biodiversity (Hubbell, 2005, 2001) (such as the neutral community model - NCM (Sloan et al., 2006)). According to traditional niche-based theory, microbial communities are shaped by deterministic factors such as competition and niche differentiation (Sloan et al., 2006), creating a strong link between microbial population traits and the specific environment in which they grow. The controversial neutral theories break the link between microbial population composition and environment; identifying stochastic factors such as death, birth, immigration and speciation as the main drivers for microbial assembly (Hubbell, 2005). The neutral theory has been cast as a simple mathematical model with a small number of parameters and has been able to describe the structure (species abundance distribution and species-area relationships) of a wide range of micro- and macro-communities (Volkov et al., 2007), from fish (Muneepeerakul et al., 2008) and beetles (Ulrich and Zalewski, 2007) to some natural microbial communities (Sloan et al., 2006). However, despite being successful in describing some ecological patterns, stochastic factors alone cannot always explain all the variation observed in complex microbial communities, in which case a combination of the two types of factors (stochastic and environmental) seem more effective in describing complex realities (Ofiteru et al., 2010; Stegen et al., 2012).

In this study, we investigated the development of microbial communities on two drinking water filter materials (quartz sand and granular activated carbon) receiving the same source water in laboratory-scale columns. The NCM was used as a null model and employed to test how much of the variation in the filter microbial communities could be described by neutral stochastic processes of random deaths, births within the local community and immigration from the same metacommunity. If neutral processes were predominant in the assembly of lab-scale filter communities a good agreement between the model predictions and the experimental data would be found. On the other hand, any deviations from the model assumptions (expressed by a poor agreement) would represent the predominance of deterministic forces over stochastic processes, induced by the filtration process itself or by the presence of two contrasting filtering materials. In addition, we assessed what effect the different filter media, and their respective communities, had on the function of the water filters including DOC removal, potential THM formation and putative pathogen removal.

## 2. Methods

### 2.1. Columns set-up and operation

The biological community developing on two different filter materials, quartz sand (Sigma Aldrich, Germany) and granular activated carbon (GAC) (activated charcoal, untreated granular - Sigma Aldrich, Germany), was studied by employing a suite of four lab-scale columns, two replicates for each filter media. The filters consisted of stainless steel 30 cm columns with 0.6 cm inner diameter. Columns were packed with clean and autoclaved filter material (3.6 g of GAC and 15 g of sand) to obtain a filter depth of 27 cm (Fig. 1). Particle size ranged from 0.25 mm to 0.84 mm for GAC, and from 0.21 mm to 0.29 mm for sand. The columns were operated for 56 days (5280 empty bed volumes) at room temperature (21.9 ± 2 °C). According to previous literature, this period of time was considered sufficient to ensure the development of a stable biofilm through the whole length of the filter as suggested by the steady-state removal of DOC observed (Velten et al., 2011). The feed water was collected weekly from a local boating pond in Newcastle upon Tyne (UK) and stored in the dark at 4 °C prior to use, for no more than seven days after collection. The water was pre-filtered through cellulose paper filters (Whatman® 10 μm pore diameter, UK) and pumped through the columns with a peristaltic pump (Watson Marlow) in down flow mode. Columns were operated to mimic a rapid gravity filter and the flow rate was controlled to an average of 0.42 mL/min which corresponded to a hydraulic loading of 1 m/h and an average empty bed contact time of 15.5 min.

### 2.2. Sampling and chemical analyses

Influent and effluent waters were regularly sampled to monitor chemical and microbiological parameters. Samples were collected after 0, 6, 19, 26, 34, 43, 56 days of operation and directly analysed for anions (ion exchange chromatograph Dionex ICS-1000, Sunnyvale, CA), dissolved organic and inorganic carbon (Shimadzu 5050A, Japan) and ammonium ($NH_4^+$) on a Spectroquant® Pharo 300 spectrophotometer (Merck, USA) using Spectroquant® test kits (Merck, USA).
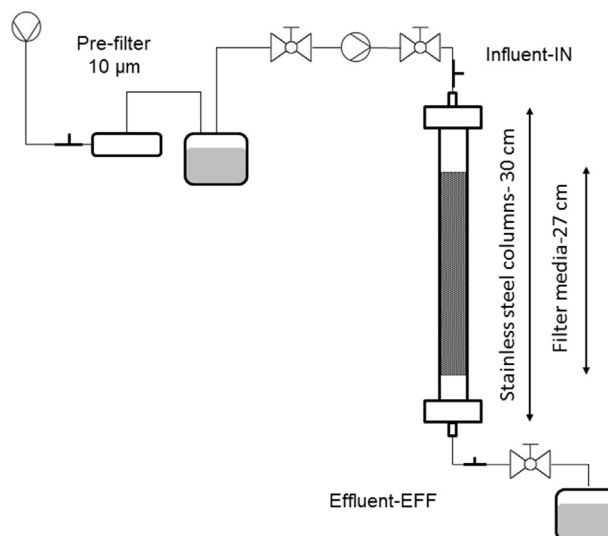


**Fig. 1.** Schematic presentation of the filter columns.

THM formation potential was assessed on the effluent water collected after 50 days of reactor operation. Sodium hypochlorite was added to the water to a final concentration of 100 mg/L of chlorine; THMs were measured with a GC-ECD (Agilent, CA) (Werner et al., 2016) after 1, 6 and 23 h of contact time following a protocol of the University of Massachusetts (Reckhow, 2006). The amount of THMs formed was normalised by the amount of DOC initially present in the water.

## 2.3. Total cell concentration

The total cell concentration in water samples (for both influent and effluent) was analysed by fluorescence staining of nucleic acids combined with quantitative flow cytometry (Hammes et al., 2008), using a FACScan flow cytometer (Becton Dickinson, NJ) equipped with a 15-mW 488-nm air-cooled argon-ion laser. Water samples were collected and fixed in absolute ethanol (1:1 v/v), stored in the dark at 4 °C and analysed within 16 h. Cells were stained with 10 $\mu$L mL$^{-1}$ SYBR® Green I (10,000× in DMSO, Sigma Aldrich, Germany) diluted 1:100 in TE-buffer (10 mM Tris—HCl 1 mM EDTA; pH 8.0) and incubated in the dark for 13 min before measurement. Where necessary, samples were diluted in filtered (0.22 $\mu$m; polyethersulfone membrane, Merck Millipore, NJ) DI water, so that the events per second detected by the instrument were always less than 400. Readings were collected in logarithmic mode and analysed with Flowing Software 2.0; electronic gating was used to separate signals from background (Hammes et al., 2008; Hammes and Egli, 2010).

## 2.4. Microbial community characterisation

Three sets of samples were collected for molecular analysis: samples were collected in duplicate from the influent (IN) and effluent (EFF) waters at day 0 and after 56 days (T0; T56); and filter media samples (from sand, herein referred to as SAND in tables and figures, and GAC columns) were collected in duplicate at the end of the experiment, after the columns were decommissioned, from the top (TOP, 0—10 cm) and from the bottom (BOT, 17—27 cm) of the columns. Water samples were filtered through sterile 0.2 $\mu$m PES membrane filters (PALL, NY); 150 mL of the influent and 250 mL of the effluent. Different volumes of water were filtered for influent and effluent samples in order to obtain similar concentrations of DNA after the extraction, due to the reduction of suspended cells through the filter. The filter membranes with collected biomass were transferred to sterile Petri dishes and stored at −20 °C until extraction. Filter media samples were collected and stored in a solution of sterile PBS:Ethanol 50% v/v at −20 °C. DNA was extracted from the membrane filters containing the biomass, and from 0.5 g of the column filter material using the FastDNA® SPIN Kit for Soil (MP-Biomedicals, USA) following the manufacturer's protocol, but adding three more cycles of homogenisation in the FastPrep® instrument and incubating (55 °C for 5 min) the elute before final centrifugation.

Barcoded amplicon libraries were generated by PCR amplifying the V4 and V5 regions of the bacterial and archaeal 16S rRNA genes, using primers 515F (5′-GTGNCAGCMGCCGCGGTAA-3′) and 926R (5′—CCGYCAATTYMTTTRAGTTT-3′). All forward primers contained a sequencing adaptor, a 'GT' spacer and a unique 12 base pair Golay barcode in order to allow multiplexed analyses. PCR amplifications were conducted using the FastStart High Fidelity PCR System and the PCR Nucleotide Mix (Roche Diagnostics GmbH, Germany). The following thermocycler program was used for the PCR reaction: a denaturation cycle at 95 °C for 4 min followed by 25 cycles of denaturation at 95 °C for 1 min, annealing at 55 °C for 45 s, and elongation at 72 °C for 1 min, and with a final elongation step at

72 °C of 7 min. Amplification products were purified using Agencourt AMPure XP reagent (Beckman Coulter, CA) and quantified using a Qubit dsDNA HS Assay kit (Invitrogen, CA) on a Qubit® 2.0 Fluorometer, following the manufacturer's protocols. The single amplicon libraries were pooled in equimolar amounts, further purified using a Pippin Prep System (Life Technologies, CA) following the manufacturer's protocol, and sequenced on an Ion Torrent Personal Genome Machine System (Life Technology, CA) on an Ion 316 chip according to the manufacturer's protocols.

All data processing was conducted using QIIME 1.7 bioinformatics pipeline (Caporaso et al., 2010). Briefly the sequences were trimmed to remove primers and barcodes, quality filtered (considering a minimum quality score of 20) and chimera checked (using ChimeraSlayer (Haas et al., 2011)). The sequences were clustered into operational taxonomic unit (OTUs) at a similarity level of 97%, using the UCLUST algorithm. Taxonomy assignment was conducted using the RDP naïve Bayesian rRNA classifier (Wang et al., 2007) with Greengenes database (http://greengenes.lbl.gov/) (DeSantis et al., 2006) Statistical analysis was run using PRIMER v.6 (Clarke and Gorley, 2006) and Minitab software (Minitab 17 Statistical Software, Minitab Inc., USA). The sequences obtained were deposited in the European Nucleotide Archive with study accession number PRJEB22254 (secondary accession number ERP103926). Microbial community composition was compared by non-metric multidimensional scaling (NMDS) , principal coordinate analysis (PCO), and cluster analysis of Bray Curtis similarities (Chase et al., 2011).

Quantitative PCR (qPCR) was used to quantify the number of bacterial cells present on the filter medium samples. The 338F (Bakke et al., 2011) and 1046R (Huber et al., 2009) primers were used to target the 16S rRNA gene, assuming that each organism in our microbial community contained one copy of the gene. qPCR assays were performed on a CFX96 real-time PCR detection system (Bio-Rad, United Kingdom), using the following temperature profile: 98 °C for 3 min for 1 cycle; and 98 °C for 5 s, followed by 60 °C for 5 s, for 39 cycles.

Each amplification reaction was run in triplicate and contained 3 $\mu$l of template DNA, to assure a concentration between 10 and 100 ng/mL of genomic DNA, (or molecular-grade water for blanks), 0.5 $\mu$l of forward and reverse primer (10 pmol/$\mu$l), 5 $\mu$l of SsoFast EvaGreen Supermix (Bio-Rad, United Kingdom), and 1 $\mu$l of molecular-grade water. Standards, in concentration ranging between $10^2$ and $10^8$ fragment copies per $\mu$l, were obtained from circular plasmids containing the target fragment of DNA, and a calibration curve was generated in every qPCR run. Melt-curve analysis (between 65 and 95 °C) was performed at the end of each qPCR run to assess the specificity of the products. The average efficiency of the qPCR analyses was 95%.

## 2.5. Ecological analyses

A modified version of the Neutral Community Model (Morris et al., 2013; Sloan et al., 2006) was used to identify OTUs present in the target communities (i.e. the local filter communities) as the effect of neutral, stochastic processes of random deaths and births within the local communities and stochastic immigration from the metacommunity. The NCM was therefore used as a null model; compliance with which would be indicative that neutral stochastic processes predominate in community assembly in lab-scale water filters, and differences from which would be indicative of deterministic processes predominating. The composition of the metacommunity was inferred by averaging the composition of the influent water samples at both sampling points T0 and T56, as proposed by Morris et al. (2013) in their application of the NCM, in order to include the highest number of filter colonisers in the

analysis. The samples employed to represent the local target communities were those collected in duplicate from the TOP and BOT portions of the four filter columns; the frequency of detection was expressed as the number of filter material samples where a specific OTU was detected over the total number of filter samples. A further analysis was conducted where the filter media samples were considered as source communities (and employed to infer the metacommunity composition) and the filter effluent samples (collected in duplicate), at T56, as the local target communities. Only the OTUs shared between source and target communities were employed for the analyses. The model uses 95% binomial confidence intervals (using the Hmsic package in R) (Morris et al., 2013) and OTUs falling between these intervals were considered to be present as a result of neutral stochastic processes of random births and deaths within the local communities and stochastic immigration from the source community (influent water or filter medium), therefore following neutrality. OTUs that fell outside the upper bound of the confidence interval were those characterised by higher frequencies of detection in the local target communities (filter medium or effluent water) than predicted by the neutral model, based on their abundance in the source. OTUs that fell outside the lower bound of the confidence interval were those detected less frequently in the local target communities than predicted by the NCM.

Any deviation from the model (upper and lower bound of the confidence interval) would represent the predominance of non-neutral processes such as environmental selection and competition and would account for those taxa that are advantaged or disadvantaged by the local community environment. The best fit of the model with the data was found by calibrating the coupled parameter $N_Tm$, where $m$ is the immigration rate and $N_T$ is the total number of individuals in the local community that was considered stable after 56 days of filter operation.

The phylogenetic community structure was also characterised using the Mean-Nearest-Taxon-Distance (MNTD) and the Nearest-Taxon-Index (NTI) as described in (Stegen et al., 2012). MNTD finds the phylogenetic distance between each OTU within a sample and its closest relative in the same sample; therefore it measures how phylogenetically close the taxa in a community are related to each other. Very low MNTD values are typical of communities composed of taxa that are highly related and, therefore, indicative of those communities where environmental selection has overcome stochastic assembly (Stegen et al., 2012). NTI quantifies the number of standard deviations that the observed community MNTD is from the mean of a null distribution. NTI greater than $+2$ indicates a community whose taxa are more closely related than expected by chance, whereas an NTI less than $-2$ are indicative of communities where coexisting taxa are more distantly related than expected by chance. This method has shown to successfully distinguish niche-based assembled local communities from neutrally assembled ones by dispersal using simulated metacommunities (Kembel, 2009). We employed this test to evaluate how much selection has impacted the assembly of our communities over and above stochastic forces. Average NTI values were assigned to SAND and GAC filters employing the duplicates samples from TOP and BOT portions.

## 3. Results and discussion

### 3.1. The effect of neutral processes on microbial assembly

We observed that the Neutral Community Model accounted for only 28% (measured as Pearson coefficient, p = $6.2 \times 10^{-11}$ confidence intervals: 0.19–1.38) of the observed variation in the frequency of OTUs as predicted by their relative abundance in the influent water. Indeed, the majority of the taxa employed in the NCM analyses (the ones shared between influent water and filter media communities), 54%, were either more frequent (24% of taxa) or less frequent (30% of taxa) than would be expected by their relative abundances if they were assembled by neutral stochastic processes alone (Table 1., Fig. 2a). This suggests that neutral forces can at best only explain a small part of the microbial assembly and that deterministic forces are predominant in the system. If the assembly of the microbial community in the two filter media were neutral, rather than determined by environmental selection in the filter, the mean relative abundance of each OTU in the influent would dictate the frequency with which that OTU is detected in the filter, regardless the filter medium type. Therefore, plotting the relative abundance of OTUs in the influent water against the detection frequency of OTUs in the filter would result in a continuous monotonically increasing curve converging to 100 (Fig. 2). Other studies have assessed the consistency of the neutral theory with the assembly of natural microbial communities. A study on the microbial composition of lung communities showed a much higher correlation between the model prediction and the empirical observations than that observed by us (Spearman rank correlation > 0.84) (Morris et al., 2013). Nevertheless, it seems that the correlation we observed is comparable to that in another engineered system (activated sludge), where the Neutral Community Model could account for 20% of the variability (measured as $R^2$) of the heterotroph community (Ofiteru et al., 2010).

Moreover, the microbial communities observed on the filter media of our lab-scale filters showed a composition of taxa more phylogenetically related than expected if they were stochastically assembled (NTI greater than $+2$). Therefore, the filtration process itself (for our specific conditions and at a laboratory scale) exerted a strong selective effect within the 56 days of our study, if we assume that all the taxa must have originated from the source community. Among the filtration processes we studied, a further level of selection was posed by the presence of two contrasting materials. It appeared that sand exerted a higher environmental pressure than GAC selecting for a phylogenetically closer community: the mean sand-NTI observed was significantly higher than the mean GAC-NTI, 8.41 and 7.16, respectively (t-Test, p = 0.013) (Table SI1).

In total 683 OTUs were detected among influent and filter media samples, 503 OTUs were detected on the filter material, while 588 OTUs were detected in the filter influent samples, 408 were shared between the source and the filter community and were employed in the model. Only 46% of the shared OTUs, equal to 37% of the total OTUs detected in the filter media, fell in the neutral region, constructed with the 95% binomial confidence intervals (Table 1, Fig. 2a). The same analysis was performed on the effluent communities, this time considering the filter material as a source of dispersal. Here, we observed that the effect of neutral stochastic processes on the community assembly was more pronounced; indeed, the model was able to describe 50% (Pearson coefficient, p = $5.7 \times 10^{-11}$, confidence interval: 0.37–0.65) of the observed variation in the community (Fig. 2b). However, 35% of taxa were either more frequent (12%) or less frequent (23%) than would be expected by their relative abundances if neutral processes alone had assembled them. The improved fit with neutrality reflects the stochastic nature of the filtration process itself; water flows through the filter sloughing off part of the community that has developed on the filter material and what is present in higher abundance on the filter is then found in higher frequency in the effluent water (Sloan et al., 2006). Among the 631 OTUs detected between filter media and effluent water, 611 were detected in the effluent and 503 on the filter media; 483 were shared between filter media and effluent water (Table 1). Almost all the OTUs present on the filter media were detected in the effluent water, except

**Table 1**
Description of the OTUs employed in the NCM analyses.

| Influent and filter media OTUs | | | Filter media and effluent OTUs | | |
|---|---|---|---|---|---|
| Influent-Source | 588 | | Filter-Source | 503 | |
| Filter-Target | 503 | | Effluent-Target | 611 | |
| Shared | 408 | | Shared | 483 | |
| Enriched | 98 | 24% | Enriched | 60 | 12% |
| Disadvantaged | 122 | 30% | Disadvantaged | 110 | 23% |
| Neutral | 188 | 46% | Neutral | 313 | 65% |
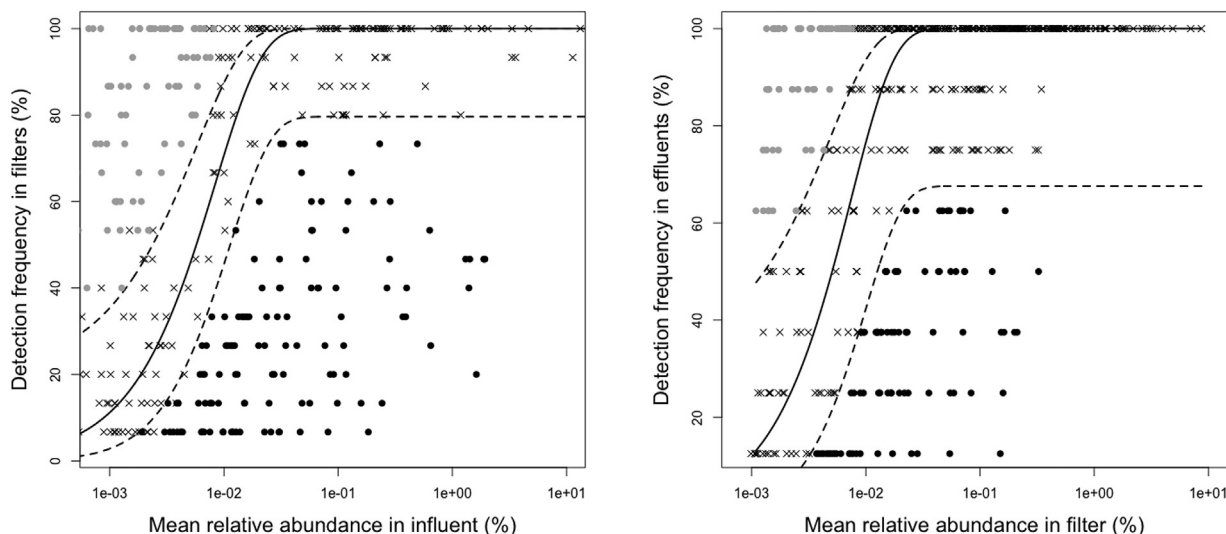


**Fig. 2.** Neutral Community Model analyses for filters (a) and for effluent samples (b). The solid line is the model prediction and the dashed line is the 95% confidence intervals. Grey circles (above the upper 95% confidence interval), represent the OTUs for which the observed frequency is greater than the model prediction (enriched) and black circles (below the lower 95% confidence interval), represent the OTUs where the frequency is less than the prediction (disadvantaged) based on their mean relative abundance in the metacommunity.

20, which confirms that the filter community seeds the effluent water (Pinto et al., 2012). However, 128 OTUs that were present in the filter effluent were not detected on the filter samples; these OTUs that were "passing through" (Pinto et al., 2012) the filter were originally present in the influent water, but were unable to colonise it. They represented 21% of the total OTUs detected in the effluent.

These findings suggest that filter communities are not strongly assembled by neutral processes and that the filtration process generates a selective environment where deterministic forces overcome stochastic ones within 56 days of operation (Dini-Andreote et al., 2015).

### 3.2. The effect of filter medium type on microbial communities in the filters

Communities that developed on the surface of the two different materials showed significant differences in their similarity (ANO-SIM, Global R = 0.976 and p<0.1%): different samples from the two replicates tended to cluster according to the material on which they developed (Fig. 3). Such clustering seems likely to be the result of the different properties of the two types of materials used, such as intraparticle porosity, surface area, chemistry and adsorption capacity. Furthermore it must be noted that the particle sizes of the two different materials, though similar, were more varied for GAC than sand. In more detail, *Planctomyces* was the most abundant genus on sand filters (SAND), representing 11% and 13% of the top (TOP) and bottom (BOT) communities followed by *Gemmata* at 8.1% (TOP) and 1.3% (BOT) and by two other taxa, whose genera were not assigned, from the *Pirellulaceae* and *Saprospiraceae* families. The two OTUs with the highest relative abundance on GAC filters were

unidentified beyond the class level: a *Gammaproteobacteria* constituted 5.6% and 6.2% of the TOP and BOT GAC communities, and an unknown bacteria constituted 5.3% and 5.2% of the two communities. Those OTUs that contributed the most to the dissimilarity between the two filter communities were identified using the SIMPER (SIMilarity PERcentages) analysis in PRIMER. These were taxa of the genus *Planctomyces* and *Gemmata* that were detected with a high relative abundance on both materials, but with higher values observed in sand rather than in GAC. In addition, species belonging to the genus of *Gordonia*, *Sulfuritalea* and
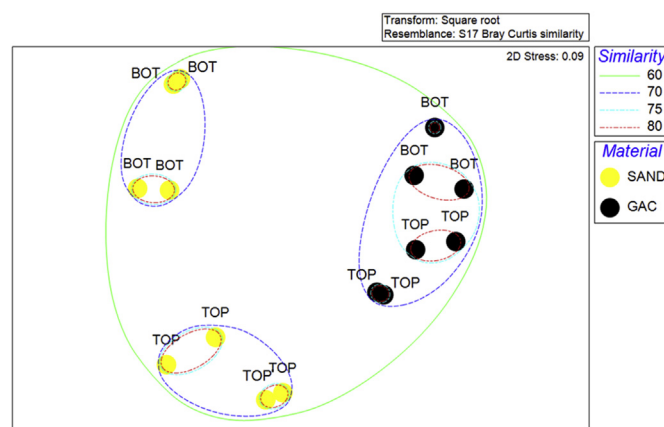


**Fig. 3.** NMDS generated using the Bray-Curtis index for the TOP and the BOT columns samples from the two different materials.

*Nitrospira* were identified as main contributors and were detected in very low abundances on GAC, but with higher abundances in sand. Two other taxa, one belonging to the genus *Hydrogenophaga* and the other to the family *Opitutaceae*, were detected mainly in GAC columns. In a previous study *Hydrogenophaga*-related bacteria were one of the most predominant bacteria isolated from different GAC filters sampled in the Netherlands (Magic-Knezev et al., 2009).

### 3.3. The effect of filter medium type on microbial communities in the effluent water

Influent and effluent water samples differed significantly over time and between filter media (Two-way crossed ANOSIM, Global R = 1 and 0.998 respectively, p < 0.001 for both factors). The influent communities, at T0 and at T56, showed 53% similarity, indicating a change of the feeding water between these sampling times (Fig. 4a).

Among the effluent samples, the bacterial community seemed to cluster according to the filter material. Samples of the two physical replicates, for each filter material, clustered together with a similarity of 85%, for both sand and GAC, at T0 and 75% at T56. PERMANOVA analyses showed that both filter influent microbial composition and filter material influenced the microbial composition of the effluent water (p < 0.001 for both factors). However, the influent water microbial composition seems to be the main contributing factor. According to Principal Component Analysis, influent water explained 60.8% of the total variation (Fig. 4b), while the filter medium type explained 14% of the effluent water composition variation in microbial community similarities. The similarity between influent and effluent water communities decreased from the beginning to the end of the experiment: from 70% at T0 to 65% at T56. The 30% dissimilarity between influent and effluent water at T0 can only be explained as a direct result of the filtration process, where microorganisms in the influent water were removed from the water during the passage through the filter, since no other microorganisms were present on the two media at the beginning of the experiment. The slight dissimilarity between the two effluent types at this stage of the experiment (15%) is probably the result of different porosity created by the two materials due to different grain structures.

At the end of the experiment, the similarity between influent and effluent decreased along with the similarity between the

effluents from the two different filter media, from 70% to 65% and from 80% to 75%, respectively (Fig. 4a). In this case, the presence of established microbial communities on the filter material might explain the increase of both dissimilarities: part of the community developed on the filter seeds the effluent water thereby affecting its composition. Indeed, some of the taxa whose concentration increased from influent to effluent water were those enriched in the filter environment.

Moreover, the most abundant bacteria detected in the influent water samples belonged to the class of *Actinobacteria* (15% and 33% at T0 and T56), *Flavobacteria* (24% and 11% at T0 and T56) and *Betaproteobacteria* (30% and 10% at T0 and T56). *Planctomycetia* (12% GAC, 26% sand), *Alphaproteobacteria* (17% GAC, 12% sand) and *Gammaproteobacteria* (9% GAC, 4% sand) were the most abundant classes found on the filter biofilm. Previous studies have observed that certain species inhabiting freshwater have higher capacity to attach to surfaces and survive in biofilms, while others mostly survive in a planktonic state (Douterelo et al., 2013). In freshwater ecosystems, *Flavobacteria* often lives in an entirely planktonic state growing on specific algal exudates (Zeder et al., 2009). This would explain the high relative abundance of those bacteria classes in the influent water and their very low abundance on biofilm (less than 1%).

### 3.4. Implications for water quality: cell removal

Suspended cells were removed from the influent water through the columns (Fig. 5a). Sand and GAC removed the same concentration of cells; no statistically significant difference was observed between the two effluents at each sampling time (Paired *t*-Test, p-value > 0.172). Both materials removed, on average 66 ± 9% of the cells entering the columns. However, the concentration of cells detected on the columns at the end of the experiment differed between the two materials. GAC hosted less cells per $cm^3$ than sand despite having a higher specific surface area and a rougher structure that would allow a better cells colonisation. In total $4.91 \pm 0.48 \times 10^9$ cells/$cm^3$ and $2.90 \pm 0.20 \times 10^9$ cells/$cm^3$ colonised sand and GAC, respectively (Fig. 5b), which is in the same range as previously measured cell concentrations in drinking water biofilters (Lautenschlager et al., 2014; Velten et al., 2007). We also calculated that during the whole duration of the experiment, $1.42 \times 10^{11}$ cells entered the columns and $4.51 \times 10^{10}$ left the system, therefore $9.66 \times 10^{10}$ cells were either retained in the columns
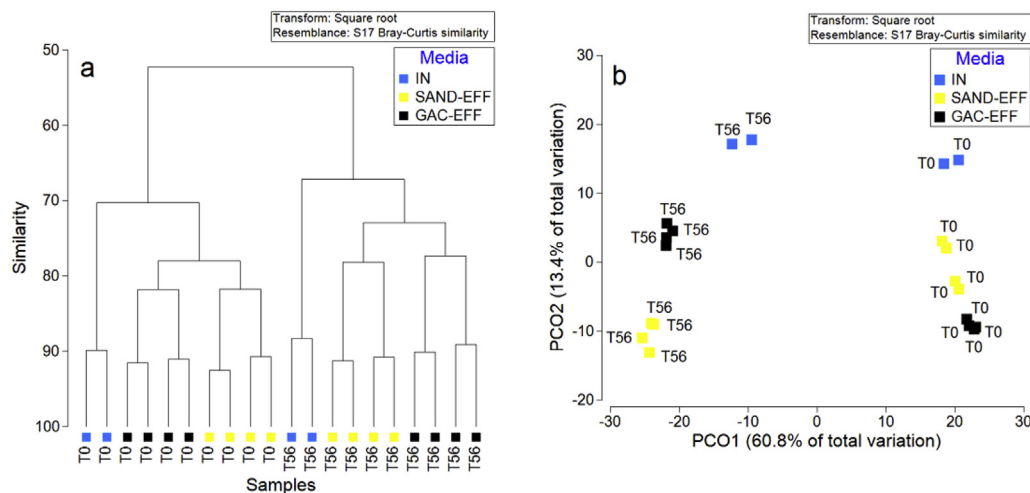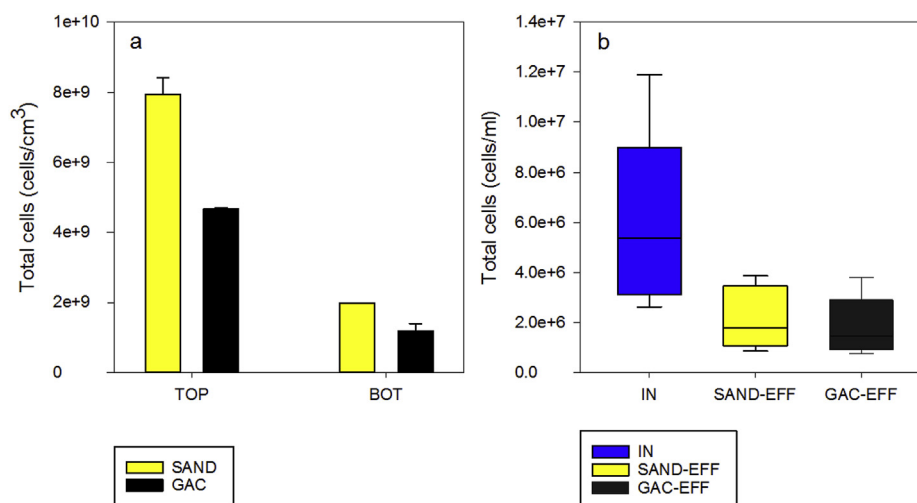


**Fig. 4.** (a) Dendrogram of the similarity index of influent (IN) and effluent water (EFF) at the two sampling points (T0, T56) and (b) PCO plots of the influent (IN) and effluent water at the two sampling points (T0,T56).

**Fig. 5.** (a) Total cells on top (TOP) and bottom (BOT) positions of the columns measured at T56. (b) Average concentration of total suspended cells in the influent (IN) and the effluent water from Sand and Granular Activated Carbon columns (SAND-EFF, GAC-EFF) throughout the experiment (T0, T6, T19, T26, T34, T43, T56) as determined by flow cytometry.

or removed from the system. A total of $3.75 \pm 0.64 \times 10^{10}$ cells colonised the sand columns at the end of the experiment, while $2.22 \pm 0.30 \times 10^{10}$ cells colonised the GAC columns, suggesting that in both materials, cell decay is an important removal mechanism. This net decay was higher in GAC (a minimum of $7.74 \times 10^9$ cells, ignoring cell growth on DOC) as compared to sand (a minimum of $5.62 \times 10^9$ cells). This analysis demonstrates that cell removal from the influent water is not merely due to physicochemical entrapment and screening out of cells by the filter material, but also involves cell lysis or predation by heterotrophic nano-flagellates. Note that the cell decay could in reality be even higher than calculated, since we did not consider cell growth in our cell number balance.

The GAC adsorption capacity was very high at the beginning of the experiment; in fact almost 40% of the DOC present in the water was removed, reducing the amount of DOC available in the water for microbial growth. This might lead to higher cell decay in the filters and to the lower colonisation of GAC.

On both materials, more cells were measured on the top of the columns rather than the bottom (Fig. 5b). This was most likely due to higher readily biodegradable substrate availability in this region, a stratification already observed in previous studies (Lautenschlager et al., 2014; Velten et al., 2011), and from the decay of cells filtered onto the columns from the influent. The higher amount of cells detected on the top positions of the filters can be linked with the higher observed alpha biodiversity. In fact, while, no differences in terms of alpha biodiversity were observed between the two materials, on both materials, a higher diversity was observed at the top positions (Table SI2 of supplementary information).

### 3.5. Implications for water quality: putative pathogens and faecal indicators

Six main genera that harbour putative pathogens and faecal indicators were identified in the influent and effluent water collected at T56 and on the filter material: *Mycobacterium*, *Aeromonas*, *Clostridium*, *Legionella*, *Enterobacter* and *Flavobacterium*; from here on described as putative pathogens.

The relative abundances of those genera were multiplied by the total cell concentrations measured using flow cytometry (expressed in cells/mL) or qPCR, in order to estimate the final total abundances (Fig. 6a, Fig. 6b).

The concentrations of those genera harbouring putative pathogens in the influent water were significantly higher than the concentrations in the effluent water (*t*-test, p < 0.02), except for *Legionella*, whose concentration increased from influent to effluent (*t*-test, p < 0.006), and was found significantly higher in sand effluent (*t*-test, p < 0.007). The concentration of *Mycobacterium*, *Clostridium* and *Flavobacterium* cells were significantly lower in sand effluent (*t*-test, p < 0.04) than in GAC effluent, and no significant difference was observed between the two effluent types for *Enterobacter* cells. Only *Mycobacterium*, *Clostridium*, *Legionella* and *Flavobacterium* were detected on the filter media. *Mycobacterium* cells were found at significantly higher abundance on sand columns in both top and bottom positions (*t*-test, p < 0.02). *Clostridium* cells were detected only in the top positions of the columns in both materials, but no significant difference was detected between the two media; *Legionella* cells were detected on both media in both positions with a significant higher abundance of cells on the bottom position of sand columns (*t*-test, p < 0.015). Finally, *Flavobacterium* cells were detected on both materials in both positions, but no significant difference was detected between the two materials.

The absence (or their presence at abundances lower than detection limits) of *Aeromonas* and *Enterobacter* cells on filter material agrees with previous studies showing poor survival of those bacteria in drinking water biofilm (Kwon et al., 2011; Lehtola et al., 2007). The filtration process is confirmed as an effective treatment for ensuring a two-log reduction of faecal indicators from the influent water over 27 cm total filter depth, regardless of the material employed. To explain the observed two-log reduction between influent and effluent for these "passing through" bacteria, which were absent from the filter biofilm community, we estimated very high decay rates, equal to $12.0 \text{ h}^{-1}$ for *Aeromonas* in both filter materials and $8.11 \text{ h}^{-1}$ and $11.7 \text{ h}^{-1}$ for *Enterobacter* in sand and GAC respectively.

While *Aeromonas* and *Enterobacter* cells seem to act as taxa that "pass through and/or decay", the rest of the putative pathogens identified colonised the filter. In the case of *Legionella*, this led to a net cell growth with significant release of cells into the effluent water, resulting in increased number of *Legionella* cells compared to the influent. Species of *Mycobacterium* and *Legionella* have been
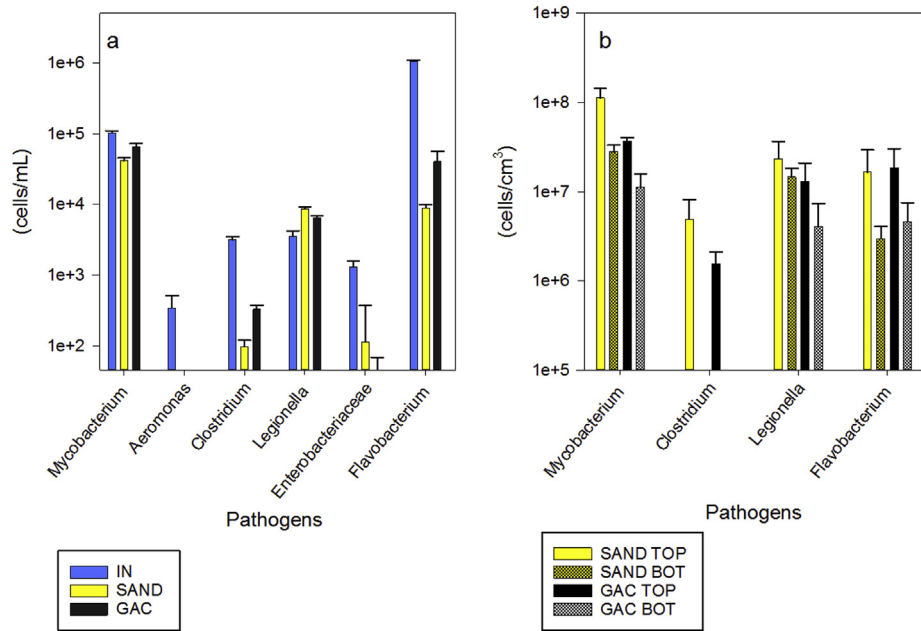
**Fig. 6.** (a) Putative pathogen concentrations in the influent and effluent water at T56 and on (b) the filter.

detected in several drinking water biofilms and a previous study has found that these bacteria are able to accumulate in biofilms even under high-shear turbulent-flow conditions (Lehtola et al., 2007). The distinct behaviour of different genera harbouring putative pathogens in the water filtration process, which tends to shape communities in distribution networks, demonstrates a need for more comprehensive characterisation and monitoring of the drinking water microbiome.

It must also be acknowledged that further research would be necessary to confirm the identity of the taxa identified in this study (and referred to as "putative pathogens") as true pathogens and check their viability.

### 3.6. Implications for chemical water quality: DOC removal and THM formation potential

DOC removal differed between GAC and sand packed columns
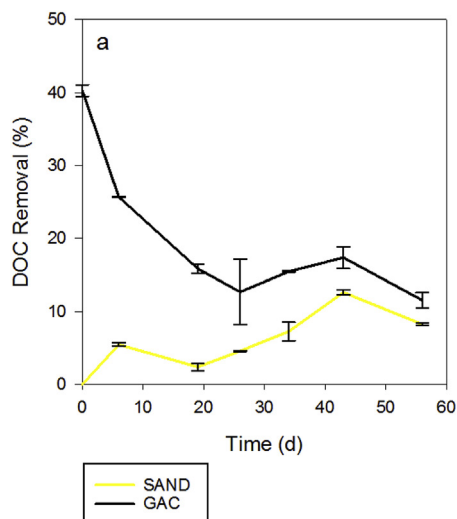
(Fig. 7a). In GAC columns, removal started high at 40%, due to DOC adsorption, but dropped after 19 days stabilising at an average of 15%. The stable DOC removal from day 25 onwards is most likely due to biological activity compared to the chemical/physical adsorption that occurred in the first 25 days. In sand columns, no significant DOC removal was recorded in the first 20 days, while it progressively increased reaching an average of 9% in the following 36 days, which, as in the GAC columns, could likely be attributed to biology (Velten et al., 2011).

The difference in DOC removal efficiency was also reflected in the THM formation potential of the filter effluents, which was 277.9 ± 16.3 μg THMs/L for sand and 188.8 ± 14.1 μg THMs/L for the GAC columns (at 23 h chlorine contact time). The organic matter present in water is known to be the main contributor to the formation of several classes of undesirable disinfection by-products, where THMs are the most prevalent. The THM formation potential (THMFP) increases with the concentration of DOC present in
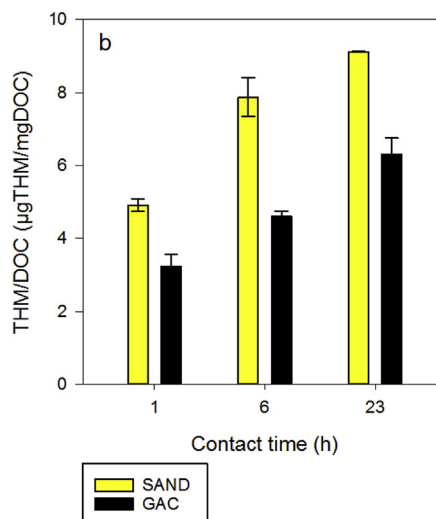


**Fig. 7.** (a) DOC removal rate for GAC and Sand water filters over time (b) THMFP of SAND and GAC effluent at different contact times.

water (Ramavandi et al., 2015).

It was therefore not unexpected that a higher concentration of THMs would occur in the sand column effluent, where the DOC removal was slightly less effective. However, the amount of THMs normalised to the quantity of DOC present in the water was, at any of the three different chlorine contact times, still much higher for sand column effluent than for GAC (Fig. 7b).

It can be inferred that the two materials preferentially removed different fractions of DOC, with GAC being more effective in removing the DOC fractions that are responsible for the highest amount of THM produced. It has been demonstrated that the chemical characteristics of dissolved organic matter influences the formation of disinfection by-products: specific DOC fractions were found to have a higher THMFP than others (Ramavandi et al., 2015; Yang et al., 2008; Zhang et al., 2009). Since DOC is partially biodegradable, selective DOC removal by the two filter media would also create two different environmental conditions for the filter colonizing microorganisms. Whether the preferential removal of THM precursors in GAC columns was only due to chemical adsorption or to the combination of chemical activity and biological degradation was not evaluated. Further studies are required to assess reliably how the different microbial communities of the two filter media distinctly degrade such compounds.

## 4. Conclusions

- This is the first study that investigates the role played by neutral and deterministic factors in the assembly of filter microbial communities in a controlled setting.
- The assembly of lab-scale water filter communities is only weakly influenced by neutral processes: the filtration process itself exerts an environmental pressure and, between the two materials investigated, sand selected for communities more phylogenetically closely-related than GAC.
- Two different communities developed on the two materials triggered a difference between the two effluent communities and affecting the overall quality of the effluent water. Higher abundance of *Legionella* was found on the sand grains resulting in a significantly higher concentration of *Legionella* cells in the sand effluent compared to GAC. This suggests that such material could preferentially be used to control *Legionella* released in the distribution network.
- The different putative pathogens showed very distinct behaviours, suggesting that traditional faecal indicator bacteria, such as *E. coli* , may not adequately represent other pathogens such as *Mycobacterium avium* and *Legionella pneumophila* in water treatment.
- Overall this study has showed that the filter medium is a deterministic factor that can drive the assembly of different microbial communities resulting in the improvement or deterioration of effluent water quality. This raises the exciting opportunity that such filter communities could be purposefully engineered to improve drinking water quality.

under an 'Open Data Commons Open Database License'. Additional metadata are available at: https://doi.org/10.17634/101925-1. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.watres.2017.09.042.

## References

Bakke, I., De Schryver, P., Boon, N., Vadstein, O., 2011. PCR-based community structure studies of Bacteria associated with eukaryotic organisms: a simple PCR strategy to avoid co-amplification of eukaryotic DNA. J. Microbiol. Methods 84, 349–351. https://doi.org/10.1016/j.mimet.2010.12.015.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303.

Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., Inouye, B.D., 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2. https://doi.org/10.1890/ES10-00117.1 art24.

Clarke, K., Gorley, R., 2006. PRIMER V6: User Manual/Tutorial.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05.

Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., Salles, J.F., 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proc. Natl. Acad. Sci. U. S. A. 112, E1326–E1332. https://doi.org/10.1073/pnas.1414261112.

Douterelo, I., Sharpe, R.L.L., Boxall, J.B.B., 2013. Influence of hydraulic regimes on bacterial community structure and composition in an experimental drinking water distribution system. Water Res. 47, 503–516. https://doi.org/10.1016/j.watres.2012.09.053.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methé, B., DeSantis, T.Z., Consortium, T.H.M., Petrosino, J.F., Knight, R., Birren, B.W., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504. https://doi.org/10.1101/GR.112730.110.

Hammes, F., Berney, M., Wang, Y.Y., Vital, M., Koster, O., Egli, T., 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. Water Res. 42, 269–277. https://doi.org/10.1016/j.watres.2007.07.009.

Hammes, F., Egli, T., 2010. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. Anal. Bioanal. Chem. 397, 1083–1095. https://doi.org/10.1007/s00216-010-3646-3.

Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Funct. Ecol. 19, 166–172. https://doi.org/10.1111/j.0269-8463.2005.00965.x.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

Huber, J.A., Morrison, H.G., Huse, S.M., Neal, P.R., Sogin, M.L., Mark Welch, D.B., 2009. Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. Environ. Microbiol. 11, 1292–1302. https://doi.org/10.1111/j.1462-2920.2008.01857.x.

Kembel, S.W., 2009. Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. Ecol. Lett. 12, 949–960. https://doi.org/10.1111/j.1461-0248.2009.01354.x.

Kwon, S., Moon, E., Kim, T.-S.S., Hong, S., Park, H.-D.D., 2011. Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. Microbes Environ. 26, 149–155. https://doi.org/10.1264/jsme2.ME10205.

Lautenschlager, K., Hwang, C., Ling, F., Liu, W.-T., Boon, N., Köster, O., Egli, T., Hammes, F., 2014. Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. Water Res. 62, 40–52. https://doi.org/10.1016/j.watres.2014.05.035.

Lautenschlager, K., Hwang, C., Liu, W.T., Boon, N., Köster, O., Vrouwenvelder, H., Egli, T., Hammes, F., 2013. A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. Water Res. 47, 3015–3025. https://doi.org/10.1016/j.watres.2013.03.002.

Lehtola, M.J., Torvinen, E., Kusnetsov, J., Pitkänen, T., Maunula, L., Von Bonsdorff, C.-H.H., Martikainen, P.J., Wilks, S.A., Keevil, C.W., Miettinen, I.T., 2007. Survival of Mycobacterium avium, Legionella pneumophila, Escherichia coli, and Calicivirases in drinking water-associated biofilms grown under high-shear turbulent flow. Appl. Environ. Microbiol. 73, 2854–2859. https://doi.org/10.1128/AEM.02916-06.

Magic-Knezev, A., Wullings, B., Van der Kooij, D., 2009. *Polaromonas* and *Hydrogenophaga* species are the predominant bacteria cultured from granular activated carbon filters in water treatment. J. Appl. Microbiol. 107, 1457–1467. https://doi.org/10.1111/j.1365-2672.2009.04337.x.

Morris, A., Beck, J.M., Schloss, P.D., Campbell, T.B., Crothers, K., Curtis, J.L., Flores, S.C., Fontenot, A.P., Ghedin, E., Huang, L., Jablonski, K., Kleerup, E., Lynch, S.V., Sodergren, E., Twigg, H., Young, V.B., Bassis, C.M., Venkataraman, A., Schmidt, T.M., Weinstock, G.M., 2013. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am. J. Respir. Crit. Care Med. 187, 1067–1075. https://doi.org/10.1164/rccm.201210-1913OC.

Muneepeerakul, R., Bertuzzo, E., Lynch, H.J., Fagan, W.F., Rinaldo, A., Rodriguez-Iturbe, I., 2008. Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. Nature 453, 220–222. https://doi.org/10.1038/nature06813.

Ofiteru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., Sloan, W.T., 2010. Combined niche and neutral effects in a microbial wastewater treatment community. Proc. Natl. Acad. Sci. U. S. A. 107, 15345–15350. https://doi.org/10.1073/pnas.1000604107.

Pinto, A.J., Xi, C.W., Raskin, L., 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environ. Sci. Technol. 46, 8851–8859. https://doi.org/10.1021/Es302042t.

Ramavandi, B., Farjadfard, S., Ardjmand, M., Dobaradaran, S., 2015. Effect of water quality and operational parameters on trihalomethanes formation potential in Dez River water, Iran. Water Resour. Industry 11, 1–12. https://doi.org/10.1016/j.wri.2015.03.002.

Ramette, A., Tiedje, J.M., 2007. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proc. Natl. Acad. Sci. U. S. A. 104, 2761–2766. https://doi.org/10.1073/pnas.0610671104.

Reckhow, D.A., 2006. Analysis of Trihalomethanes and Related Pentane-extractable Organic Halides. University of Massachusetts, pp. 1–33.

Richter, D., Massmann, G., Dünnbier, U., 2008. Behaviour and biodegradation of sulfonamides (p-TSA, o-TSA, BSA) during drinking water treatment. Chemosphere 71, 1574–1581. https://doi.org/10.1016/j.chemosphere.2007.11.026.

Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., Curtis, T.P., 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ. Microbiol. 8, 732–740. https://doi.org/10.1111/j.1462-2920.2005.00956.x.

Stegen, J.C., Lin, X., Konopka, A.E., Fredrickson, J.K., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J. 6, 1653–1664. https://doi.org/10.1038/ismej.2012.22.

Tilman, D., 1982. Resource competition and community structure. Monogr. Popul. Biol. 17, 1–296.

Ulrich, W., Zalewski, M., 2007. Are ground beetles neutral? Basic Appl. Ecol. 8, 411–420. https://doi.org/10.1016/j.baae.2006.08.002.

Velten, S., Boller, M., Koster, O., Helbing, J., Weilenmann, H.U., Hammes, F., 2011. Development of biomass in a drinking water granular active carbon (GAC) filter. Water Res. 45, 6347–6354. https://doi.org/10.1016/j.watres.2011.09.017.

Velten, S., Hammes, F., Boller, M., Egli, T., 2007. Rapid and direct estimation of active biomass on granular activated carbon through adenosine tri-phosphate (ATP) determination. Water Res. 41, 1973–1983. https://doi.org/10.1016/j.watres.2007.01.021.

Vital, M., Dignum, M., Magic-Knezev, A., Ross, P., Rietveld, L., Hammes, F., 2012. Flow cytometry and adenosine tri-phosphate analysis: alternative possibilities to evaluate major bacteriological changes in drinking water treatment and distribution systems. Water Res. 46, 4665–4676. https://doi.org/10.1016/j.watres.2012.06.010.

Volkov, I., Banavar, J.R., Hubbell, S.P., Maritan, A., 2007. Patterns of relative species abundance in rainforests and coral reefs. Nature 450, 45–49. https://doi.org/10.1038/nature06197.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07.

Werner, D., Valdivia-Garcia, M., Weir, P., Haffey, M., 2016. Trihalomethanes formation in point of use surface water disinfection with chlorine or chlorine dioxide tablets. Water Environ. J. 30, 271–277. https://doi.org/10.1111/wej.12209.

Yang, X., Shang, C., Lee, W., Westerhoff, P., Fan, C., 2008. Correlations between organic matter properties and DBP formation during chloramination. Water Res. 42, 2329–2339. https://doi.org/10.1016/j.watres.2007.12.021.

Zeder, M., Peter, S., Shabarova, T., Pernthaler, J., 2009. A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. Environ. Microbiol. 11, 2676–2686. https://doi.org/10.1111/j.1462-2920.2009.01994.x.

Zhang, H., Qu, J., Liu, H., Wei, D., 2009. Characterization of dissolved organic matter fractions and its relationship with the disinfection by-product formation. J. Environ. Sci. 21, 54–61. https://doi.org/10.1016/S1001-0742(09)60011-5.