Diversity and Dynamics of the Microbial Population Associated with Drinking Water Distribution Systems and their Impact on Drinking Water Quality

Report to the
WATER RESEARCH COMMISSION

by

S Vosloo[1], M Crous[1], K Moodely[1], L Gounden[1], S MacRae[1], M Sigudu[2], A Pinto[3] and SN Venter[1]

[1]Microbiology and Plant Pathology, University of Pretoria,
[2]Process Technology, Rand Water, Vereeniging,
[3]Civil and Environmental Engineering, Northeastern University, Boston, USA

WRC Report No. 2469/1/18
ISBN 978-1-4312-0016-3

August 2018

Obtainable from
Water Research Commission
Private Bag X03
GEZINA, 0031

orders@wrc.org.za or download from www.wrc.org.za

This report is supplemented with a CD at the back containing electronic copies of all figures in the report as well as supplementary data.

Printed in the Republic of South Africa

EXECUTIVE SUMMARY

BACKGROUND

Water distribution systems are complex aquatic environments with a high diversity of microorganisms, which could have a major impact on the microbial quality of the drinking water supplied to consumers. Most of the microbes present in the water have typically survived the treatment processes and are present in the final water at low levels. Microbial regrowth is believed to be responsible for most of the problems associated with drinking water distribution systems and could impact on the water quality both in terms of safety and aesthetics (Boe-Hansen *et al*., 2002). Without a clear understanding of the diversity of microorganisms present and the dynamics within this ecosystem, management of the microbial quality can only be done on an ad hoc basis when problems are experienced.

The recent development in high throughput sequencing, e.g. Illumina (Hall, 2007), has opened the door to investigate microbial communities at species level resolution as well as the ability to quantify the members of the community. This approach has improved our understanding of the ecology, diversity, interactions and functioning of diverse microbial ecosystems, such as the human gastrointestinal system (Human Microbiome). These techniques also have the potential to be used for assessing the microbial community within drinking water distribution systems.

RATIONALE

Using this approach, Pinto and co-workers collected data on the bacterial community of a drinking water treatment plant and distribution system over a year (Pinto *et al*., 2012 and 2014). These researchers were able to demonstrate that the community in the distribution system was closely linked to the community associated with the filters and that the raw water only had a limited impact. They also showed that temporal changes in the community could be closely linked to seasonal changes. These researchers predicted that this type of community analysis approach could form the basis of a predictive framework for the management of microbial quality and risks associated with drinking water systems (Pinto *et al*., 2014).

Water distribution systems have their own unique characteristics and it was uncertain whether the findings of Pinto and co-workers would be valid for the large distribution systems present in South Africa. During the current study we therefore addressed the following questions based on data collected from a South African system:

- What is the variation in the microbial community present in a distribution system primarily supplied by treated surface water and can the variation be linked to season factors?

- Is the same correlation between the microbial community in the bulk water and the sand filter community observed in a large distribution system where different disinfection regimes are practiced at different times?
- How homogenous is the microbial community associated with sand filters?
- Are the bacteria involved in nitrification already present in the bulk water directly after treatment and what factors promote their growth within the system?

OBJECTIVES AND AIMS

The overall objective of this study was to collect information and data that would provide the necessary foundation for the development of a predictive framework and early warning system to manage the microbial quality of drinking water supplied to consumers. A better understanding of the microbial ecology of the water distribution system could also be used to optimize sampling strategies. The possibility of using this information to ultimately manipulate the drinking water microbiome has also been raised (Pinto *et al*., 2014).

The project aims were:

1. To determine the dynamics of the microbial community present in a large chloraminated water distribution system.
2. To correlate potential population shifts in the microbial community with changes in water quality parameters and environmental conditions associated with the distribution system.
3. To determine the impact of the biofilm community associated with the rapid sand filters on the microbial community present in the treated drinking water within the distribution system.
4. To determine the homogeneity of the microbial community associated with rapid sand filters within two large drinking water treatment works.
5. To determine the presence of potential opportunistic pathogens that may pose a risk to consumers.

APPROACH

The main focus of the project was to address the abovementioned objectives by sampling a large treatment and distribution system, representative of the systems operated by the larger utilities in the country. The project however also included samples from three other systems or subsystems. They included 1) a smaller treatment and distribution system, supplying chlorinated surface water, 2) a small system supplying groundwater after limited treatment and 3) a municipal reticulation system supplying water to consumers after it has obtained the water from one of the water utilities. Although all these systems were from the same geographical area, it would be possible to apply the main findings of the project to other drinking water utilities in South Africa as the treatment as well as distribution conditions and management practices are fairly representative of South African systems.

Sampling was done on a monthly basis for all the drinking water treatment and distribution systems investigated. Biomass was collected either through concentration of the bulk water samples or sonication and filtration of the sand samples collected from the filter beds. After DNA extraction, the genomic DNA was sent to the University of Michigan, Medical School (Ann Arbor, United States of America) for sequencing of the V4 hypervariable region of the 16S rRNA gene, using the Illumina MiSeq platform with a pair-ended sequencing protocol as described by Kozich *et al*. (2013). The sequencing data were processed according to the standard approached used for 16S profiling, using the MOTHUR platform and its associated bioinformatics tools. Alpha and beta diversity indexes were calculated for all communities sampled and were used to investigate spatial and temporal variation within the different treatment and distribution systems under investigation.

RESULTS AND DISCUSSION

Aims:
To determine the dynamics of the microbial community present in a large chloraminated drinking water distribution system.
To correlate potential population shifts in the microbial community with changes in water quality parameters and environmental conditions associated with the distribution system.
To determine the impact of the biofilm community associated with the rapid sand filters on the microbial community present in the treated drinking water within the distribution system.

Contrary to what was initially expected based on the literature, it was observed that in the large distribution system, the community present on the sand filter media was not the main driver in shaping the microbiome present the DWDS. This study clearly demonstrated that the microbial communities inhabiting a large drinking water distribution system were strongly shaped by spatial rather than temporal factors (Chapter 4). Although other physiochemical and environmental factors had some impact, it was clear that chlorination and subsequent chloramination were the main drivers shaping the microbial community in this system. This was not unexpected, as it is known that different disinfection processes will have a significant impact on the microbial community.

The study also investigated whether the molecular approach utilised during the main project (Chapter 3 and 4), could be applicable to and benefit South African communities supplied by smaller drinking water and reticulation systems. From the data presented in Chapter 5, it was clear that the approach to study the microbial ecology of water treatment and distribution systems using 16S profiling is also applicable to alternative systems. This approach provided valuable information on the impact of treatment, distribution, seasons and other water quality parameters on the final microbial quality of the water supplied to the consumer, especially when alternative treatment approaches or water sources were used.

Aim:

To determine the homogeneity of the microbial community associated with rapid sand filters within two large drinking water treatment works.

Based on the sampling performed at two different treatment works (Chapter 3), it was established that although some level of local and intra variation amongst the rapid sand filter microbial communities might exist, no significant spatial differences were detected amongst sand filter microbial communities (i.e. along the surface and depth of the RS filter bed, and across parallel RS filters within a filter gallery) sampled during the same sampling run. These findings were based on the comparison of beta diversity indexes. The data clearly indicated that the sand filter community is homogenously distributed across the filter bed at a specific point in time as well as between filters within the same filter house. This has definite implications for sampling efforts as it implies that the dominant members of the filter community within a filter house could be detected based on a single representative sample taken from one of the filters in the filter house.

Aim:

To determine the presence of potential opportunistic pathogens that may pose a risk to consumers.

One of the main benefits of the 16S profiling approach is that it provides a detailed inventory of the bacterial species or rather operational taxonomic units (OTUs) present in a water sample. It also provides data on their relative abundance at the time of sampling. This data was used to investigate the presence of potential bacterial pathogens (waterborne and water-based/opportunistic) in the drinking water system. Out of the 8891 OTUs initially observed in the dataset only 18 could potentially be linked to pathogens previously detected in drinking water (Chapter 6). Of these 14 were closely related to water-based bacterial pathogens such as members of the families *Sphingomonadaceae, Aeromonadaceae*, and *Staphylococcaceae*. Four of the OTUs were closely related to waterborne bacterial pathogens such as *E. coli* and *Enterococcus* spp. None of the opportunistic pathogens typically associated with drinking water distribution systems and with the potential to cause outbreaks, such as *Legionella pneumophila* and *Mycobacterium avium complex,* were detected in the data analysed during this study.

Apart from the scientific outputs, this project also contributed towards capacity building in the field of water microbiology. Several students were involved in this project and have obtained or will obtain post-graduate degrees as a result of their involvement in this project. A full list of list of the capacity building achieved as well as conference presentations are provided at the back of the report.

CONCLUSIONS

This study has clearly shown that drinking water treatment and distribution systems could differ markedly from each other and that a universal model to predict the microbial community of the water supplied to the consumer would be difficult to achieve. The current study demonstrated that the necessary technologies and

knowledge is available to study the microbial ecology and dynamics of individual treatment and distribution systems. Understanding the ecology and the factors that shape the drinking water microbiome is essential when appropriate measures to manage the microbial quality and associated health risks of drinking water in such a system, are to be developed and implemented.

This study again confirmed that groundwater, from a well-protected aquifer, remains a valuable source of drinking water, as microbes are present at very low levels. It was shown that for such systems only minimal disinfection and treatment are required to ensure safe drinking water. Attention should however, be given to the distribution of this water as no residual disinfectants are present to counter contamination events.

The health risk posed by the potential bacterial pathogens detected in the distribution system investigated are currently not of great concern within the distribution system as they were mostly detected sporadically at generally low abundances in comparison to the total drinking water microbial community. The study did not focus on premise plumbing systems where monitoring for some of these opportunistic pathogens may be of value as the conditions within the plumbing system may differ marked from those associated with the distribution and reticulation system. The presence of these bacteria in the distribution system during this study was possibly due to their survival or growth in the system. Many of these water-based species are ubiquitous and are often detected in environmental samples including water. The distribution patterns observed were not indicative of either treatment failure or a specific contamination event.

RECOMMENDATION FOR FUTURE RESEARCH

This study clearly showed that each step within the distribution and reticulation system had an impact on the quality of the water that reaches the consumer. The main focus of the present study was on the distribution system and did not address the effects of the reticulation system on the microbial water quality in great detail. In reticulation systems, issues such as the impact of smaller diameter pipes as well as retention times within the smaller community reservoirs on the bacterial community are important. These issues have not been addressed in detail during this study and it is recommended that future studies focusing specifically on the microbial ecology and dynamics of the reticulation and premise plumbing systems should be undertaken.

The present study has not focused on the specific functions performed by the community in the system. This could be addressed by metagenome studies, which will assist in clarifying the metabolic processes performed by the community. Furthermore, genomic binning of metagenomic data can be applied to investigate individual genomes of the dominant microorganisms in the system. This approach will clearly assist in understanding the biological interactions in the systems in greater detail and create valuable information, which could be used to manage the biostability of drinking water distribution systems.

ACKNOWLEDGEMENTS

The authors would like to thank the Reference Group of the WRC Project for their assistance and the constructive discussions throughout the duration of the project:

Dr N Kalebaila (Chairperson)	Water Research Commission
Dr TG Barnard	University of Johannesburg
Mr I Bailey	Umgeni Water
Prof C Bezuidenhout	North West University
Mr J Kutu	City of Tshwane
Mr W Le Roux	CSIR
Ms A Maurizi	City of Ekurhuleni
Mr S Nduli	Magalies Water
Mr V Ntema	Magalies Water
Mr M Vulindlu	City of Cape Town
Prof G Wolfaardt	Stellenbosch University

The project team would like to thank all the water utilities and municipalities who participated in this project. A special word of thanks to the specific individuals who assisted with the necessary permission, sampling logistics and collection of additional data.

TABLE OF CONTENTS

All the Appendixes form part of a separate electronic file containing the supplementary data. This file is provided as part of the report but will not be printed.

LIST OF FIGURES

Please note that electronic copies of all of the figures in the report are included in the CD containing all the supplementary data.

LIST OF TABLES

ACRONYMS & ABBREVIATIONS

AOA	Ammonia oxidising archaea
AOB	Ammonia oxidising bacteria
AOC	Assimilable organic carbon
ATP	Adenosine triphosphate
BDOC	Biodegradable dissolved organic carbon
BFP	Biofilm formation potential
BOM	Biodegradable organic matter
CA	Chloramination
DBP	Disinfection by-products
DGGE	Denaturing gradient gel electrophoresis
DOC	Dissolved organic carbon
DON	Dissolved organic nitrogen
DOP	Dissolved organic phosphorous
DWDS	Drinking water distribution system
DWTP	Drinking water treatment plant
EPS	Extracellular polymeric substances
FB	Filter bed
FCM	Flow cytometry
FIV	Filter-independent variable OTUs
GAC	Granular activated carbon filters
HAA	Haloacetic acids
HAN	Haloacetonitriles
HNM	Halonitromethanes
HPC	Heterotrophic plate counts
LC	Leaky colonisers
MAP	Microbially available phosphorus
MIC	Microbial mediated corrosion
MRA	Mean relative abundance
NGS	Next generation sequencing
NOB	Nitrite oxidising bacteria
NOM	Natural organic matter
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PE	Plastic polyethylene
PLFA	Phospholipid fatty acid analysis
POC	Particulate organic carbon

PS Purification station
PT Pass-through OTUs
PVC Polyvinyl chloride
RGS Rapid gravity sand filters
RS Rapid sand filters, also
RSF Rapid sand filters
SC Strict colonisers
SS Suspended solids
SSCP Single-strand conformation polymorphism
SSF Slow sand filters
TCC Total cell counts
THM Trihalomethanes
TOC Total organic carbon
T-RFLP Terminal restriction fragment length

CHAPTER 1: INTRODUCTION AND OBJECTIVES

1.1 BACKGROUND

Water utilities face major challenges in supplying and maintaining safe drinking water to communities. Water leaving the treatment plant is typically of high quality but conditions within the distribution system often lead to deterioration in water quality. The treated water enters the distribution system with a microbial load, a physical particle load as well as a nutrient load. Due to the interaction of these components and varying environmental conditions within the system, water of poorer quality is often found at the consumer tap compared to the drinking water leaving the treatment plant (Lui *et al*., 2013). Amongst the quality determinants, the microbial quality of the water is typically impacted the most.

Water distribution systems are complex aquatic environments with a high diversity of microorganisms, including fungi (Siqueira and Lima, 2013), which could have a major impact on the microbial quality of the drinking water supplied to consumers. Microorganisms potentially enter the distribution system in two ways: They may enter into the distribution system from external sources such as open reservoirs, pipe breakages as wells as intrusions during maintenance of pipelines, valves or reservoirs or an increase of microorganisms in the system may be due to microbial re-growth within the systems. These organisms have typically survived the treatment processes and are present in the final water at low levels. Microbial growth is believed to be responsible for most of the problems associated with drinking water distribution systems and could impact on the water quality both in terms of safety and aesthetics (Boe-Hansen *et al*., 2002). Without a clear understanding of the diversity of microorganisms present and the dynamics within this ecosystem, management of the microbial quality can only be done on an ad hoc basis when problems are experienced.

Initial studies to investigate the microbial community of drinking water distribution and reticulation systems have strongly focused on the selective culturing of the bacteria present (Geldreich, 1996). It is now widely accepted that such studies have focused on only a limited part of the microbial community, and that both the dominant members and some important pathogens may not have been detected (Kunin *et al*., 2008). With the advent of molecular methods (PCR and sequencing), study approaches have changed however, they still focused on targeting individual species or limited representatives of the bacterial community, and data on the overall bacterial community was still sorely lacking. These molecular techniques included detection and quantification of specific species or pathotypes by PCR, fluorescent in situ hybridisation (FISH), and characterizing bacterial gene amplicons by terminal restriction fragment length polymorphism (T-RFLP). A drawback of these approaches is that the researcher often preselects the bacteria that will be detected and a comprehensive catalogue of microbial species and pathotypes present are not obtained.

The development in high throughput sequencing, e.g. 454/Roche or Illumina (Hall, 2007), and microarray technologies for community analyses, e.g. PhyloChip (Brodie *et al*., 2006 and 2007) during the last decade has opened the door to investigate microbial communities at species level resolution as well as quantitatively (Huber *et al.,* 2007; Kunin *et al*., 2008). Major advances in our understanding of the ecology, diversity, interactions and functioning of diverse microbial ecosystems, such as marine and human gastrointestinal systems, have since been made. These techniques also have the potential to be used for assessing the microbial community within drinking water distribution systems. For the first time there are molecular tools available to the researcher to generate quantitative data on the diversity of microbial communities.

Several studies have made use of high throughput sequencing for microbial community and diversity analyses of both bulk water and biofilms in distribution systems (Hong *et al.*, 2010; Kwon *et al.,* 2011; Hwang *et al*., 2012b; Gomez-Alvarez *et al*., 2012; Pinto *et al*., 2012; Zhang *et al*., 2012; Chaio *et al*., 2014; Douterelo *et al*., 2014; Sun *et al*., 2014). All these studies have demonstrated the ability of this approach to characterise the microbial communities within these systems and revealed how disinfectant regimes and various environmental factors influenced and changed the community composition and diversity. According to these studies the bacterial groups that existed within the system were highly dependent on multiple factors such as source water, disinfectant strategy (Hwang *et al*., 2012a) and pipe materials.

Using a similar approach, Pinto and co-workers collected data on the bacterial community of a drinking water treatment plant and distribution system over a year (Pinto *et al*., 2012 and 2014). These researchers demonstrated that the community in the distribution system was closely linked to community associated with the filters and that the raw water only had a limited impact. They also showed that temporal changes in the community could be closely linked to seasonal changes. These researchers predicted that this type of community analysis approach could form the basis of a predictive framework for the management of microbial quality and risks associated with drinking water systems (Pinto *et al*., 2014).

Water distribution systems have their own unique characteristics and it is uncertain whether the findings of Pinto and co-workers will be valid for the large distribution systems present in South Africa. During the proposed study we would therefore like to address the following questions based on data collected from a South African system:

- What is the variation in the microbial community present in a distribution system primarily supplied by treated surface water and can the variation be linked to season factors?
- Is the same correlation between the microbial community in the bulk water and the sand filter community observed in a large distribution system where different disinfection regimes are practiced at different times?
- How homogenous is the microbial community associated with sand filters?
- Are the bacteria involved in nitrification already present in the bulk water directly after treatment and what factors promote their growth within the system?

It was anticipated that answers to these questions will provide the necessary foundation for the development of a predictive framework and early warning system to manage the microbial quality of drinking water supplied to consumers.

1.2 PROJECT AIMS AND OBJECTIVES

The overall aim of the current study was to collect the information and data that could provide the necessary foundation for the development of a predictive framework and early warning system to manage the microbial quality of drinking water supplied to consumers. A better understanding of the microbial ecology of the water distribution system could also be used to optimize sampling strategies. The possibility of using this information to ultimately manipulate the drinking water microbiome was also raised (Pinto *et al*., 2014).

The project objectives were:

1. To determine the dynamics of the microbial community present in a large chloraminated water distribution system.
2. To correlate potential population shifts in the microbial community with changes in water quality parameters and environmental conditions associated with the distribution system.
3. To determine the impact of the biofilm community associated with the rapid sand filters on the microbial community present in the treated drinking water within the distribution system.
4. To determine the homogeneity of the microbial community associated with rapid sand filters within two large drinking water treatment works.
5. To evaluate the potential risk to consumers associated with the microbial population present in the water distribution system. The presence of opportunistic pathogens that may pose a risk to immune compromised patients will also be evaluated.

1.3 SCOPE AND LIMITATIONS

The main focus of the project involved the treatment and distribution systems operated by a large water utility. The project also analysed three other systems or subsystems (reticulation systems). They included a small system supplying chlorinated surface water, a small system supplying groundwater after limited treatment and a municipal reticulation system supplying water obtained from a large water utility were studied. Although all these systems were from the same geographical area, it would be possible to apply the main findings of the project to other drinking water utilities in South Africa as the treatment as well as distribution conditions and management practices are fairly representative of South African systems.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION INTO MICROBIAL ECOLOGY

Drinking water treatment and distribution systems are recognised as complex aquatic environments offering multiple unique habitats that support microbial growth through the different stages of treatment and distribution. Drinking water leaving the treatment plant may be of high biological quality however, the treated water is subjected to various conditions within the distribution system, which may ultimately lead to the deterioration of water quality at the consumers' tap. Maintaining the biologically stability of the drinking water is one of the biggest obstacles facing drinking water utilities. Lautenschlager *et al.* (2013) and Prest *et al.* (2016), described biological stability as the condition where the microbial community within a drinking water distribution system (DWDS) does not change in its composition and concentration over time and space, i.e. from the treatment plant, throughout the distribution system and to the consumer's tap. This is often not possible to achieve as the microbial community can change dramatically from treatment to tap. Drinking water at the tap can contain up to 10^6-10^8 microbial cells per litre (Hammes *et al*., 2008), comprising of a highly diverse microbial community, including bacteria, archaea, eukaryota and viruses (You *et al*., 2009; Thomas and Ashbolt, 2011; Siqueira and Lima, 2013; Lui *et al.,* 2013a; Pinto *et al*., 2014; Gall *et al*., 2015).

Biological instability, i.e. changes in the bacterial community, can only be managed based on a clear understanding of the microbial ecology within the treatment and distribution system. The microbial ecology of DWDSs is governed by multiple environmental and engineering factors as well as operational conditions that influence the composition and structure of microbial communities present in the bulk water, biofilms, and sediments (Wang *et al*., 2014; El-Chakhtoura *et al*., 2015). The study of microbial ecology aims to determine what mechanisms govern the assembly of microbial communities, and what drives the changes in community composition and structure along different spatial and temporal scales (Hanson *et al*., 2012, Nemergut *et al*., 2013). Understanding the microbial ecology will help answer questions about the community's origin, stability and the interactions that restrict them or keep them stable (Gülay *et al*., 2016).

Insight into the composition and structure of microbial communities of different environmental settings has led to the better understanding of species abundance, distribution and biogeographical patterns. The mechanisms that govern species abundance and biogeographical patterns within microbial communities is not uniform across different environments. However, microbial community assemblages may be governed by four ecological processes i.e. drift, dispersal, speciation, and selection. These ecological processes rarely operate in isolation, although the relative importance of each process may vary across different taxonomic groups and environmental settings (Nemergut *et al*., 2013). Speciation and dispersal may introduce new species into communities, whereas drift and selection affect change in the presence, absence and relative abundance of species within communities (Roughgarden, 2009; Vellend, 2010).

Microbial communities are not limited to a single taxonomic group but are typically highly diverse with a high level of species present at low relative abundances and fewer species present at high relative abundances (Ferrenberg *et al*., 2013). Although these low-abundance species may increase the diversity of the community, they may not have a significant impact on the dynamics of the community as a whole in an environmental setting. However, they have been shown to be important indicator of environmental disturbances (Szabó *et al*., 2007; Hubert *et al*., 2009; Gülay *et al*., 2016). In microbial ecology, diversity is typically defined in terms of alpha and beta diversity. The inventory or alpha diversity, describes the diversity found within an environment on the smallest spatial scale (within a sampling point). In contrast, beta diversity, also referred to as differentiation diversity, measures the diversity between environments (amongst sampling point) (Nemergut *et al.*, 2013). Both alpha and beta diversity is defined by species richness and species abundance information (Little *et al.*, 2012) using different diversity metrics and indices.

Many alpha and beta diversity studies have focussed on the high-abundance species constituting the core microbial community, which are often linked to microbial-mediated processes (Richardson and Simpson, 2011; Albers *et al*., 2015; Gülay *et al*., 2016). Microbial-mediated processes within drinking water treatment and distribution systems not only play a beneficial role through biofiltration, but they may also have a negative impact on water quality as they may contribute to microbially induced corrosion of the DWS infrastructure and cause nitrification (Pinto *et al*., 2012; Camper, 2013; Roeselers *et al*., 2015). Furthermore, microbial communities may also harbour potential pathogens, specifically those communities associated with biofilms. Therefore, some of the major challenges facing water utilities is to minimise these negative aspects and supply microbially safe drinking water to communities. The responsibility of drinking water treatment plant (DWTP) and the drinking water distribution system (DWDS) is to limit microbial growth and to maintain and deliver microbially stable drinking water to consumers. It is important that potential water-borne pathogens are removed and the control of water quality is maintained throughout the distribution system ensuring water at the tap is safe.

Supplying drinking water that is both chemically and microbially safe as well as acceptable in terms of taste, odour and appearance is vital to public health and economic growth (Geldreich, 1996; Hunter *et al.,* 2010). Therefore, understanding the microbial ecology, that is, the abundance and distribution of drinking water microbial communities as well as their interactions with the environment and each other will improve microbial management strategies for drinking water industries (Bautista-de los Santos *et al*., 2016). If the ecology of the drinking water microbiome is fully understood, potential predictive modelling can be formulated and implemented. This will enable potential manipulation of conditions within the DWDS to obtain desired results (Little *et al.*, 2012; Liu *et al.*, 2013a; Zarraonaindia *et al.*, 2013).

2.2 CURRENT UNDERSTANDING OF DRINKING WATER TREATMENT AND DISTRIBUTION SYSTEMS

Drinking water is one of the most highly monitored and regulated resources. To produce microbially safe drinking water, utilities use various treatment processes to appropriately treat the contaminants found in the source water (Hwang *et al*., 2012b). Regardless of these efforts, microbial communities persist in the distribution system. Drinking water microbial communities migrate through the DWTP, into the DWDS and into the built environment and consumer homes (Pinto *et al*., 2014). The conditions of both treatment and distribution significantly alter the microbial community (Prest *et al*., 2016).

2.2.1 Drinking Water Treatment

System design of conventional treatment system

In the attempt to control microbial growth within DWDSs, a combination of treatment/purification strategies are often applied. Treatment processes typically involve the sequential use of processes such as coagulation, flocculation and sedimentation, filtration and disinfection (Hwang *et al*., 2012b). These treatment operations function independently by utilizing specific chemical and physical means to produce drinking water that is acceptable aesthetically (i.e. acceptable in colour, odour, and taste), hygienically (i.e. free of opportunistic and true pathogenic microbial species) and operationally (i.e. limited biocorrosion, fouling of pipes or nitrification) (Pinto *et al*., 2012; Zeng *et al*., 2013; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Li *et al*., 2017). The selection of treatment strategy is a fundamental decision as it influences the concentration of organic and inorganic nutrients as well as the microbial community composition and abundance within the DWDS (Prest *et al*., 2016). Furthermore, the choice of treatment strategy is highly site specific as it selected based on the characteristics of the source water, which can be highly diverse.

Source water

Source water is obtained either from surface water systems (i.e. rivers, dams, and lakes) or groundwater systems (i.e. aquifers) and is continuously delivered to DWTPs where it is subjected to sequential treatment operations (Pinto *et al*., 2012; Zeng *et al*., 2013; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Albers *et al*., 2015; Gülay *et al*., 2016; Li *et al*., 2017). Groundwater is generally stable in its composition and is typically an anaerobic, oligotrophic environment, where organic nutrients are present at very low levels. The microbial communities in these environments are usually present in very low numbers and exist in a starvation state (Morita, 1997; Prest *et al*., 2016). Alternatively, surface waters usually contain high levels of organics and bacterial cell numbers and may require additional treatment strategies. The chemical and microbial composition of surface water is frequently changing due to seasonal variations i.e. dry seasons or high rain fall periods both effecting the treatment (Van der Bruggen and Vandecasteele, 2003). Surface waters require more extensive pretreatment than groundwaters due to their high organic contents. In the case of surface waters with a high levels of natural organic matter (NOM), processes such as coagulation, flocculation and sedimentation may need to be applied (Edzwald, 1993; Matilainen *et al*., 2010).

Coagulation, flocculation and sedimentation

Typically, coagulation, flocculation, and sedimentation are used in combination with sand filtration to remove large particles (settled and non-settled particulates) and decrease turbidity, depending on the composition of the source water. Settled particulates, which are larger and denser particulates are easily removed through gravitational settling or flotation, whereas non-settled particles, that are generally smaller in size, are not as easily removed (Levine *et al*., 1991; Van Nieuwenhuijzen *et al*., 2004; Metsämuuronen *et al*., 2014). The removal of non-settled particulates are often facilitated by the addition of chemical coagulants. These coagulants function by destabilizing the negative electrostatic surface charges of the non-settled particulates, allowing them to aggregate into larger and denser particulates that are more easily removed. The water is directed towards the flocculation chambers where it is gently mixed to allow for the collision and aggregation of the destabilized non-settled particulates, which are later removed in the sedimentation chambers (Menezes *et al*., 1996; O'Melia *et al*., 1997).

Rapid gravity sand filtration

Although the initial physicochemical treatment operations (i.e. coagulation, flocculation and sedimentation) removes most of the settled and non-settled particulates, some of the smaller non-settled particulates are only removed, or reduced to lower levels, during sand filtration (Liu *et al*., 2012a; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b). Biological filtration methods typically used either rapid gravity sand filters (RGS) or slow sand filters (SSF) (Lautenschlager *et al*., 2014). Currently, different RGS filter systems exist but vary in their media configurations, which are supported on gravel at the bottom. These RGS filter systems include rapid sand (RS) filters (i.e. sand (mono media RS filters); sand and anthracite coal (dual media RS filters); sand, anthracite coal and garnet (multimedia RS filters)) and granular activated carbon (GAC) filters (Edzwald, 2010; Bitton, 2014).

Typically filters are designed to facilitate the removal of NOM and also other remnant abiotic and biotic particulates to deliver water with reduced turbidity that are usually below 0.50 NTU (Liu *et al*., 2012a; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b). Filters aim to remove bacterial growth substrates thereby limiting bacterial growth downstream in the distribution system (Pinto *et al*., 2012). For example, the levels of NOM, including dissolved organic matter (i.e. dissolved organic carbon (DOC), dissolved organic nitrogen (DON) and dissolved organic phosphorous (DOP)) and undissolved organic matter (i.e. particulate organic carbon (POC)) are removed or reduced to lower levels during RGS filtration (Liu *et al*., 2012a; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b).

The accumulation of the NOM particulates, and other abiotic and biotic particulates along the vertical depth of the RGS filter bed causes the RGS filter to clog after a period (Liu *et al*., 2012a). Clogged RGS filters are intermittently cleaned by backwashing (Edzwald, 2010; Bitton, 2014). During backwashing the flow of the water is reversed and forced up through the RGS filter bed, which resuspends the media and dislodges the entrapped particulates (Liu *et al*., 2012a; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b).

Disinfection

Finally, bacterial growth within the distribution system is then limited and inactivated through the application of disinfection processes (e.g. ozonation, the addition of chlorine and chloramination and UV treatment) (Norton and LeChevallier, 1997; von Gunten, 2003; Hwang *et al*., 2012b; Camper, 2014). Disinfection allows for the inactivation of true pathogenic and opportunistic pathogenic microbial species (Wang *et al*., 2013; Lin *et al*., 2014) and the introduction of disinfectant residuals that is key for the maintenance of biological stable water (Prest *et al*., 2016). The concept, biological stable intuitively implies that the concentration and composition of the microbial community remain unchanged during distribution (Lautenschlager *et al*., 2013). In a disinfected drinking water distribution system (DWDS), biological instability is often noted as the persistent decline in concentration of the disinfectant residuals with distribution (Gillespie *et al*., 2014; Nescerecka *et al*., 2014). This decline, if not corrected, can result in excessive microbial growth that may lead to problems such as visual deterioration of water quality (i.e. taste, odour, turbidity and discolouration) (Vreeburg and Boxall, 2007), the proliferation of pathogenic microbial species (Vital *et al*., 2012) and microbial-induced pipe corrosion (Cerrato *et al*., 2010; Sun *et al*., 2014).

2.2.2 Microbial communities in drinking water treatment plants

Microbial communities associated with source water

Water obtained from surface or groundwater systems harbour complex microbial communities that are continuously introduced into DWTPs. Previous community composition studies of bacterial communities have shown that DWTPs receiving surface water, groundwater or surface water and groundwater blends harbour complex microbial communities, comprising up to 30 phyla and in excess 3 000 OTUs (Kwon *et al*., 2011; Pinto *et al*., 2012; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Albers *et al*., 2015; Gülay *et al*., 2016; Li *et al*., 2017). DWTPs are typically dominated by *Proteobacteria*, representing more than 50% of the entire bacterial community. However, in addition to *Proteobacteria*, the presence and abundances of other phyla including *Acidobacteria*, *Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria*, *Firmicutes, Gemmatimonadetes, Nitrospirae, Planctomycetes, Verrucomicrobia* and candidates phyla TM7 and OD1 vary throughout different DWTPs (Kwon *et al*., 2011; Pinto *et al*., 2012; Zeng *et al*., 2013; Lin *et al*., 2014; Li *et al*., 2017).

The structure and composition of the microbial community is influenced by changes in the physiochemical properties of the water as a consequence of different treatment operations (Pinto *et al*., 2012; Zeng *et al*., 2013; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Albers *et al*., 2015; Gülay *et al*., 2016; Li *et al*., 2017). The ability of bacterial species to persist in the treated water is not clear. However, previous studies have suggested that the bacterial species capable of evading treatment operations are often small in size and are independent inhabitants of suspended solids (i.e. flocs) or RGS filtration systems that are persistently introduced into the surrounding environment and downstream distribution (Aizenberg-Gershtein *et al*., 2012; Pinto *et al*., 2012; Lautenschlager *et al*., 2013; Zeng *et al*., 2013; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Liao *et al*., 2015a; Li *et al*., 2017).

Microbial communities associated with rapid gravity sand filters

In the treatment of surface waters, biofiltration is typically included in treatment operations. Previous studies have shown that the DWDS microbial community is greatly influenced by treatment operations with particular emphasis on the filtration system (Pinto *et al*., 2012; Zeng *et al*., 2013; Lautenschlager *et al*., 2014; Gülay *et al*., 2016; Li *et al*., 2017). The biofilms that develop around the granular particles making up the filter bed media contain a highly diverse microbial community. These biofilms within the filter beds maintain a stable microbial community from which selected filter-independent microbial species are constantly seeded into the downstream treatment operations and distribution system.

In their study, Pinto and colleagues (2012) identified four groups of microorganisms associated with the filter beds. They defined the first group as leaky colonisers (LC) which are the microorganisms detected in both the filter and post-filtration samples. The second group was defined as the strict colonisers (SC) and are those microorganisms associated with only the filter bed. These strict colonisers are not associated with post-filtration samples but may be to some extent, with the pre-filtration samples. The third group are defined as the pass-through (PT) individuals, which are not detected on the filter beds but are in both the pre- and post-filtration samples. Finally, the fourth group, were defined as filter-independent variable (FIV) individuals, which are also not detected on the filters, but are only found at a single sampling location.

Furthermore, in their study, Pinto *et al.* (2012) also reported that 93% of the filter media population consists of bacterial members from the LC group. Most of the LC, SC, and PT groups were comprised of members from the *Proteobacteria* phylum, where the classes *Alpha*- and *Betaproteobacteria* dominated in all three groups. The most dominant orders in the LC group were, in descending order, *Rhizobiales, Rhodospirillales, Sphingomonadales* and *Burkholderiales*. *Actinobacteria and Bacteroidetes* are also two phyla found to be dominant in the SC group and also the *Bacteroidetes* in the PT group. Based on these findings, the control and management of drinking water quality in downstream distribution could be linked to the microbial community on the sand filter beds, specifically the LC group.

The microbial community within RGS filters are involved in microbial-mediated processes. In addition to the microbial-mediated removal of ammonium or ammonia through nitrification, other hazardous chemical composites, including iron, methane, arsenic and aromatic compounds are also removed by the indigenous bacterial species of RGS filters (Albers *et al*., 2015; Bai *et al*., 2013; De Vet *et al*., 2009; Gülay *et al*., 2016). For example, species of the genus *Gallionella* are involved in iron oxidation, whereas species of the genera *Hyphomicrobium* and *Pedomicrobium* are involved in manganese oxidation, and methanotrophic species of the family *Methylococcaceae* are involved in methane oxidation (De Vet *et al*., 2009; Albers *et al*., 2015; Gülay *et al*., 2016). Moreover, using a metagenomic approach Bai *et al*. (2013) identified key metabolic enzymes involved in the oxidation of aromatic compounds (i.e. aromatic oxygenase and aromatic dehydrogenase) and arsenic (i.e. arsenite oxidase) that are presumably affiliated with the phyla *Proteobacteria, Nitrospirae*, *Chloroflexi* and *Crenarchaeota*.

2.3 DRINKING WATER DISTRIBUTION SYSTEMS

Microorganisms originate from the source water and survive initial treatment processes then enter and move through the distribution system. The microbial populations have adapted to survive the oligotrophic environment of DWDSs and microbial activity typically occurs on the pipe surfaces, in the bulk water and in loose deposits (Batté *et al*., 2003; Lui *et al*., 2014). The treated water enters the distribution system with a microbial load, a physical particle load as well as a nutrient load (although reduced compared to the source water) and as a result, as the water moves throughout the DWDS the water at the consumer tap may be of less quality than the treated water immediately leaving the treatment plant, all contributing to microbial growth within the DWDS (Lui *et al.,* 2013a). Following all attempts to remove and limit microbial growth through multiple treatment processes, microbial communities still persist under extreme conditions of low nutrient levels and disinfectant residuals (Pinto *et al.,* 2014).

2.3.1 System design of conventional distribution systems

DWDS are designed and built to act as a protective barrier to prevent the growth of microorganisms and contamination as the water travels to the consumer (Bautista-de los Santos, *et al.,* 2016). The movement of large numbers of microbial communities through the system is an integral component of the drinking water and maintaining and managing the distribution system is a highly complicated process but vital in delivering safe drinking water to the consumer (Bautista-de los Santos, *et al.,* 2016).

There are many factors influencing the system and ultimately affecting the water quality. The nature of the system itself is a challenge. Water is moved between reservoirs, through various pipelines, differing in composition, and eventually through the plumbing systems of consumer homes. Reservoirs are often points of stagnation where residence times can be long, where microbial growth potentially increases and the community structure and composition is altered. In addition, premise plumbing systems in buildings and homes can be a significant source of contamination as these systems introduce possible dead ends and a variety of attachment sites for microbial growth (Flemming *et al*., 2014).

2.3.2 Microbial communities in drinking water distribution systems

Water distribution systems are complex closed aquatic environments with a high diversity of microbial communities. Water leaving the treatment plant may be of high biological quality however, the treated water is subjected to conditions within the distribution system leading to a deterioration of water quality. Drinking water leaving the consumer tap may contain 10^6-10^8 microbial cells per litre (Hammes *et al*., 2008).

Microbial growth is responsible for many of the problems originating in the drinking water distribution systems where water quality is affected both hygienically and aesthetically (Boe-Hansen *et al.,* 2002). Microorganisms potentially enter the distribution in ways other than from the source or the filter bed media as suggested previously. They may also enter into the distribution system from external sources such as open reservoirs,

pipe breakages and maintenance and the reduction of flow velocity resulting in back washing. It is believed that microbial growth within the DWDS originates primarily from biofilm growth on the internal surface of the drinking water pipelines with only a small portion of the microbial community existing freely in the water phase (Flemming *et al.,* 2002; Farkas *et al.,* 2013).

Lui *et al.,* (2013a and 2014) describes the DWDS as being multi-dimensional, where there are typically 4 phases within the system that serve as available microenvironments for microbial growth. The 4 phases are defined and summarised as the pipe wall biofilm (formed on the inner surface of the pipe material), the bulk water (the water phase flowing through), suspended solids (SS) (the particular matter passing through the system) and loose deposits (particular matter settled on the bottom of the pipe) (Lui *et al*., 2014). There is little research on the contribution of SS and loose deposits but it has been suggested that they not be overlooked, as microbial growth in loose deposits may be comparable to that of the biofilms lining the pipe walls (Lui *et al.,* 2013a and 2013b).

Typically, bacteria dominating DWDS are ultraoligotrophic as they survive in environments with very low substrate concentrations. In short, the DWDS microbiome is dominated typically by the phylum *Firmicutes* as well as the phylum *Proteobacteria* specifically the classes *Alphaproteobacteria, Betaproteobacteria* and *Gammaproteobacteria* (Hong *et al.,* 2010; Kwon *et al.,* 2011; Douterelo *et al.,* 2014; Sun *et al.,* 2014). The proportion in which these bacterial groups exists within the system is highly dependent on multiple factors such source water, disinfectant strategy (Hwang *et al.,* 2012b), filtration processes (Kwon *et al.,* 2011; Pinto *et al.,* 2012) and pipe materials. Hwang *et al.* (2012b) describes a microbial community study where changes in community composition were directly related to changes in disinfectant type (chlorination and chloramination). Here they revealed that *Cyanobacteria, Methylobacteriaceae, Sphingomonadaceae* and *Xanthomonadaceae* were the more abundant in chlorinated water where as *Methylophilaceae, Methylococcaceae* and *Pseudomonadaceae* were abundant in chloraminated water.

The influence of biofilm development on inner pipe surfaces

Biofilms are present in all DWDS where the interface between water and the pipe wall serves as a site for the build-up of organic matter and cells, which lead to bacterial multiplication (Batté *et al*., 2003). A biofilm is typically described as a layer of microorganisms connected by extracellular polymeric substances (EPS) which facilitates attachment to pipe surfaces as well as protection from disinfectants (Wingender and Flemming, 2011).

As the bulk water moves through the distribution system, biofilms develop on the inner surfaces of the pipelines (Lui *et al.,* 2013). Biofilm development is a complex process. The properties of biofilms are continuously changing. Their distribution over the pipe surface is non-uniform and their structure is discontinuous and heterogeneous. In addition, they consist of a mixture of different microorganisms which activities differ depending on their position in the aggregate. Finally, the biofilm is potentially always reorganising as detachment, hydraulic forces, levels of nutrients and introduction of new members are constantly shaping their development (Batté *et al*., 2003). With this in mind, a biofilm may never reach a

steady state, as selection is always occurring and changes in the environmental conditions may favour different bacterial species (Boe-Hansen *et al.,* 2002).

Bacterial activity differs depending on the biofilm age, where the growth rate of a mature biofilm may be very different to the rate of initial colonisation (Boe-Hansen *et al.,* 2002). Investigations into the development of biofilms are challenging, as gaining access to the inner walls of the pipeline is often not possible as DWDS are closed systems. Many studies have therefore focused on short term artificial drinking water model systems (Henne *et al*., 2012). Multiple studies have been performed to measure the rate of biofilm formation. LeChevallier *et al.* (1990) indicated that a mature and stable biofilm could be reached after 2 weeks whereas Hallam *et al.* (2001) showed that a stable biofilm was formed within 21 days. However, there are many contradicting results. Boe-Hansen *et al.* (2002) indicated that a biofilm reaches a steady state within approximately 200 days. Furthermore, a long term study performed by Martiny *et al.* (2003) showed that a drinking water biofilm community needs years to stabilise. Therefore, it is apparent that the time it takes for the development of a stable biofilm varies greatly depending on the range of factors within DWDS that effect the growth of biofilms.

Biofilms become important when they affect the DWDS in certain ways: they are the main source of planktonic bacteria in the bulk water, they promote the deterioration of pipe surfaces resulting in microbial mediated corrosion (MIC), increase the disinfectant demand and promote disinfectant decay and are responsible for nitrification in chloraminated systems (Wilczak *et al.,* 1996). In addition, biofilms can harbour coliforms and potential pathogens and the proliferation of bacteria within the bulk water is thought to originate from biofilm detachment or shearing from the pipe wall (Batté *et al*., 2003; Ndiongue *et al.,* 2005; Berry *et al.,* 2006; Wingender and Flemming, 2011).

Microbial growth associated with loose deposits

Several studies have discovered that loose deposits can potentially play a significant role in contributing to the microbial community within DWDSs (Lui *et al*., 2014). Under favourable hydraulic conditions, the accumulation of loose deposits on the pipe bottom potentially serve as an attachment site for bacteria and a source of organic compounds resulting in a microenvironment for microbial growth (Prest *et al*., 2016). In addition, these loose deposits allow for the protection of bacteria from disinfectants and due to their mobility within the system the associated bacteria can easily reach the consumer tap (Lui *et al*., 2014).

The bacterial communities associated with loose deposits are both highly variable and diverse in comparison to the other phases. The community composition is occasionally similar to that of suspended solids, suggesting that loose deposits may simply be a result of sedimentation of the suspended solids. The community includes members that were both aerobic and anaerobic with many of the bacterial genera identified were involved in sulphur and nitrogen as well as iron and arsenic biogeochemical cycling (Lui *et al*., 2014). These findings correspond to the composition of elements typically found in loose deposits and thereby contribute to increases in corrosion processes (Sun *et al*., 2014; Prest *et al*., 2014).

There is a complex relationship between the bacterial communities within the biofilm, loose deposits and the bulk water, where the majority if research has concentrated on the biofilms lining the pipe walls. It was initially estimated that 95% of the microbial biomass in the DWDS is located in the biofilms surrounding the pipeline surfaces while the remaining 5% reside in the bulk water phase (Flemming *et al.,* 2002; Lui *et al.* 2013a and 2014). However, loose deposits and sediments have been overlooked due to difficulties in sampling (Lui *et al*., 2013a). A study by Lui *et al*. (2014) showed that together loose deposits and biofilms contribute 98% of bacterial cells where 60-90% were situated in the sediment phase.

Microbial growth in the bulk water

The bulk water phase is considered as the medium for the spread of microorganisms, nutrients and particles through the distribution system (Lui *et al.,* 2013a). It has long been assumed that bacteria in the bulk water originate from detachment of biofilms or re-suspension of the sediments rather than bacterial growth in the bulk water phase itself (Prest *et al*., 2016). In 1989, Van der Wende *et al.* stated that detachment of the biofilm was responsible for the planktonic cells present in the bulk water and that bacterial growth in the bulk water was negligible.

However, biofilms and loose deposits may not dominate the distribution system under all conditions as previously thought. Boe-Hansen *et al.* (2002) showed using ATP content and leucine incorporation, that bulk water bacteria had a higher growth rate than biofilm bacteria. They suggested that the growth of bulk water bacteria should not be over looked as their growth is significant and should be considered in bacterial growth models in DWDSs. In addition, Srinivasan *et al.* (2008) suggested that in parts of the distribution system where chlorine residuals were low, bulk water bacterial may dominate.

The microbial communities between biofilms and bulk water have also been shown to be distinct (Norton and LeChevallier, 2000; Henne *et al*., 2012; Lui *et al.,* 2014). A study by LeChevallier *et al.* (1998) showed that the bacteria in the bulk water were different to those attached to the pipe surface and therefore had a minor impact on the biofilms. More recently, Lui *et al*. (2014) observed differences bacterial community composition between bulk water and biofilms. Similar results were observed by Henne *et al*. (2012), where different core bacterial communities were observed for both biofilms and bulk water. From these findings, it has been suggested that the bulk water serves as a seed bank for sediments and biofilms and there after each phase develops its own bacterial community depending on the specific environmental conditions (Henne *et al*., 2012; Lui *et al*., 2014; Prest *et al*., 2016).

Furthermore, Pinto *et al*. (2012) showed that the composition of the bulk water community was consistently shaped by the filter bed and the dominant bacteria associated with it. With this in mind and considering the slow growth rate of oligotrophic bacteria, the microbial community within the filter may be more important for bulk water bacteria. Here, Pinto *et al*. (2014) suggest that the bacterial community on the filter can be used to predict the bacterial communities downstream in the distribution system.

Contradicting results may be explained by the variability in different distribution systems and the multiple factors that affect the microbial growth in both the biofilm and bulk water. The interaction between bacterial communities between biofilms, loose deposits and the bulk water is unclear. When considering the effects of hydraulic forces, the detachment of biofilms and re-suspension of sediments, they all undoubtedly contribute to bacterial community composition and cell concentrations (Prest *et al*., 2016).

2.4 FACTORS AFFECTING THE GROWTH OF THE DRINKING WATER MICROBIOME

It is important to understand the factors influencing microbial growth and the interactions of these factors with microbial processes in order to facilitate effective control of microbial growth within the DWDS. Conditions within the DWDS can significantly influence the biological stability of the drinking water. These factors include treatment operations and design, pipe material and distribution system infrastructure, water temperature, hydraulic forces, disinfectant and levels of nutrients and organic compounds. Often correlations between microbial growth and a single factor are not possible as it is typically a combination of multiple factors and their interactions that affect microbial growth (Camper *et al*., 2014).

2.4.1 Treatment operations and filter bed design

Although *Proteobacteria* is the most abundant phylum, and numerous other phyla are shared amongst RGS filters of different DWTPs, substantial differences are found at lower taxonomic classifications. For example, within *Proteobacteria* at class level, *Gammaproteobacteria* was reported to dominate RS filters in Europe (Albers *et al*., 2015; Gülay *et al*., 2016), whereas *Alphaproteobacteria* dominate GAC filters in Asia (Kwon *et al*., 2011; Bai *et al*., 2013; Lin *et al*., 2014; Kim *et al*., 2014; Li *et al*., 2017) and North America (Pinto *et al*., 2012). Reasonable explanations for these dissimilarities include differences in RGS filter design (Lautenschlager *et al*., 2014), treatment operations prior to RGS filtration (Fonseca *et al*., 2001; Lautenschlager *et al*., 2014) and selected physical and chemical water properties (Li *et al*., 2010; Pinto *et al*., 2012; Liao *et al*., 2013; Kim *et al*., 2014; Albers *et al*., 2015). Lautenschlager *et al*. (2014) found that the bacterial community structures of three different filters types, i.e. RS filters, GAC filters and SS filters were between 25% and 30% dissimilar, due to differences in filter design and treatment operations prior to each filtration step. In addition, previous studies have reported that the bacterial community structures of GAC filters notably changed with temperature fluctuations (Pinto *et al*., 2012; Kim *et al*., 2014), and altered concentrations of phosphorus (Li *et al*., 2010), organic carbon and nitrogen (Liao *et al*., 2013), and iron and methane (Albers *et al*., 2015).

Besides these factors, the bacterial community structures of RGS filters are also influenced by periodic backwashing, which can reduce the amount of attached bacterial cells on RGS filter particles between 35% and 50% (Kasuga *et al*., 2007; Liu *et al*., 2012a; Gibert *et al*., 2013; Liao *et al*., 2015b). The impact of backwashing on the bacterial community structures in RGS filters has been addressed since early 1999. Previous studies using phospholipid fatty acid analysis (PLFA) and T-RFLP indicated that backwashing with

water containing disinfectants, including chlorine or chloramine had a significant impact on the bacterial community structures in GAC filters (Moll and Summers, 1999; Kasuga *et al*., 2007). Similar, more recently, Liao *et al*. (2015b) indicated that backwashing with water containing no disinfectants had a significant impact on the bacterial community structure in GAC filters.

2.4.2 Pipeline Material and Distribution System Infrastructure

It is well known that the type of pipe material has a direct effect on the water quality and resulting microbial communities (Niquette *et al.,* 2000). Pipe materials not only affect the biofilm formation but also the microbial community composition, diversity and richness (Yu *et al.,* 2010; Douterelo *et al.,* 2014). Wide varieties of pipe materials have been used worldwide, depending on different distribution systems, the cost and availability of materials. A single distribution system may have a diverse range of pipe materials. Certain pipe materials can modify and increase decay of disinfectant residuals leading to an increase in microbial regrowth (Hallam *et al.,* 2001; Lehtola *et al.,* 2004; Yu *et al.,* 2010). Biofilms directly interact with the pipe surfaces and especially with cast iron pipes, microbial growth leads to microbial mediated corrosion (MIC) where the resulting corrosion by-products may provide a nutrient source or interact with the disinfectant residuals, thereby reducing disinfectant efficacy (Niquette *et al.,* 2000; Batté *et al*., 2003; Yu *et al.,* 2010).

Typically, pipeline materials can be broken down into three groups: cementitious, metallic (cooper and steel) and plastic (polyethylene and polyvinyl chloride) (Momba *et al.,* 2000; Yu *et al.,* 2010). However, whatever the pipes are made of, they all can become colonised by microorganisms. Studies have shown that copper pipes exhibit the lowest biofilm formation potential (BFP) when compared to plastic polyethylene (PE) and steel pipes, which led to lower microbial numbers in the water (Lehtola *et al.,* 2004; Yu *et al.,* 2010). Steel pipes have been shown to have the higher BFP than plastic or copper pipes and under microscopic observation, they were covered in large amounts of microorganisms (Niquette *et al.,* 2000; Hallam *et al.,* 2001; Yu *et al.,* 2010). In terms of biofilm formation, Niquette *et al.* (2000) recommended plastic over cement and steel pipes whereas Yu *et al.* (2010) recommends copper over plastic and steel pipes. Conversely, Wingender and Flemming, (2004) showed no significant differences in colonisation between steel, polyvinyl chloride (PVC) and PE pipes.

In addition, not only does the pipe composition influence the biofilm formation, but also pipe surface structure. The roughness of the pipe surface can affect bacterial attachment. Biofilm regrowth on pipes containing rough surfaces is greater than that of smooth surface. Metal pipes, typically cast iron and steel show greater biofilm formation than smooth surface plastic PVC pipes (Yu *et al*., 2010). Variation in biofilm formation in the different pipe materials is also heavily influenced by the source water and hydraulic forces (flow velocity) within different DWDS.

2.4.3 Level of Nutrients and Organic Matter

Heterotrophic bacteria require organic carbon for growth and therefore the presence of organic matter can be responsible for the proliferation of biofilms in DWDSs (Camper, 2014). The availability of nutrients in the form of organic matter, nitrogen and phosphorus can all influence bacterial regrowth and biofilm formation as well as promote disinfectant decay (Chandy and Angles, 2001; Chu *et al.,* 2005). Nutrients mainly enter the distribution system in the form of biodegradable organic matter (BOM). Nutrient levels are known to influence the amount of microbial growth within the distribution system, to the point where limiting the BOM in the system can be used to control bacterial growth (Volk and LeChevallier, 1999; Chandy and Angles, 2001). Biodegradable compounds typically either originate from the source water or from materials in contact with the bulk water, i.e. pipe surfaces. Organic compounds act as a carbon source for heterotrophic bacterial where oxygen is used as a hydrogen acceptor. Approximately 50% of the total organic carbon (TOC) is used in respiration and released as CO_2, where the remaining 50% is assimilated into cellular components, contribution to cell growth (Momba *et al.,* 2000).

BOM consists of a broad spectrum of different organic carbon compounds including simple organic sugars and acids as well as complex polymeric substances such as humic compounds (Prest *et al*., 2016). BOM can be broken down into two main components: dissolved organic carbon (DOC) and assimilable organic carbon (AOC) (Momba *et al.,* 2000). Only a small portion of DOC can be utilised by bacteria as an energy source (Prest *et al*., 2016). The DOC is a broad term for organic compounds from a wide variety of sources and variable compositions, a faction of which is biodegradable dissolved organic carbon (BDOC). BDOC can be defined as the portion of organic carbon metabolised and mineralised by heterotrophic bacteria (Camper *et al*., 2014). The AOC can be described as the fraction of the BOM that can be easily used and converted into cell mass and can therefore give an indication of the growth potential of the heterotrophic bacteria present in the system (Momba *et al.,* 2000; Camper *et al*., 2014).

While many studies have focused on organic carbon and its influence on microbial growth, inorganic compounds such as nitrogen, phosphorous and trace elements (iron, potassium, copper, magnesium etc.) also contribute to the growth of heterotrophic bacteria, although in smaller amounts (Prest *et al*., 2016). It has been shown that by limiting the availability of phosphorus, microbial growth can be controlled. Miettinen *et al.* (1997), showed that microbial growth is highly regulated by phosphorus as well as inorganic carbon and that phosphorus was the only inorganic element that had an effect on microbial growth. These results were confirmed by Lehtola *et al.* (2002), where microbial growth was limited when microbially available phosphorus (MAP) was low.

Furthermore, the type and concentration of organic and inorganic substrates determine the type of microorganisms present in the water (Prest *et al*., 2016). Though heterotrophic bacteria are known to dominate, the presence of autotrophic bacteria such ammonia and nitrite-oxidisers, iron-oxidisers, sulphate-reducing etc. have been observed in different drinking water environments. Examples of these include higher abundances of ammonia-oxidising bacteria such as *Nitrospira* and *Nitrosomonas* in waters rich in ammonium

or chloraminated waters (Wolfe *et al.,* 1990), and iron-oxidising bacteria such as *Gallionella* and *Sphaerotilus* associated with biocorrosion processes (Emde *et al*., 1992; Sun *et al*., 2014).

2.4.4 Water Temperature

Temperature has been shown to have a significant effect on microbial growth kinetics within the DWDS. Temperatures can fluctuate dramatically within a single DWDS on a seasonal basis. Pinto *et al*., (2014) observed temporal trends within a drinking water bacterial community which corresponded to seasonal cycles. Seasonal trends are often observed in a single distribution system where an increase water temperature, typically in the warmer months of spring and summer results in an increase in bacterial abundance and richness (Pinto *et al*., 2014). Increased water temperatures are often associated with increased bacterial abundance in drinking water. Water temperature can therefore affect the bacterial community composition as an increase in temperature may allow for a competitive advantage for specific bacterial groups. Temperatures from 15°C and above have been shown to increase the growth of nitrifying bacteria (Kirmeyer, 1995; Pintar and Slawson, 2003).

Therefore, an increase in water temperature in the summer months is often associated with an increased possibility of bacterial growth problems. In addition, the total effects of temperature may be influenced by other factors such as presence of chlorine residual, BOM and shear forces (Ollos *et al.,* 2003; Ndiongue *et al.,* 2005). Ngiongue *et al.* (2005) demonstrated that temperature appeared to have little effect on the level of biofilm formation before the addition of chlorine. However, when BOM was added this had a significant impact on bacterial numbers. When BOM is absent, temperature appeared to have no affect and shear forces seemed to play a more important role whereas when BOM was present, temperature was more important than shear (Ollos, 1998; Ollos *et al.,* 2003).

2.4.5 Forces Hydraulic

Among the many factors affecting the microbial ecology, many studies have investigated the effects of hydraulic forces on microbial growth (Percival *et al.,* 1999; Ollos *et al.,* 2003; Lehtola *et al.,* 2006). Bacteria, particles and dissolved nutrients are constantly being fed into the DWDS by water flow (Lui *et al*., 2014). Hydraulic conditions play an integral role in the interactions between the 4 phases described by Lui *et al*. (2014) (bulk water, biofilms, loose deposits and suspended solids). The bacteria associated with biofilms and loose particles can become resuspended in the bulk water phase and carried through the system to the consumers tap. Periods of low water demand may result in lower flow velocities and stagnation causing and increase residence time, sedimentation of particles and ultimately increases in microbial growth (Lui *et al.,* 2013a, b). Conversely, when water consumption is high, flow velocity increase causing increased bacterial dispersion through sediment resuspension and potential shearing of biofilms (Lehtola *et al.,* 2006; Douterelo *et al*., 2014; Prest *et al*., 2016). These alternating scenarios result in fluctuations in the abundance and composition of bacterial communities within the bulk water.

Here, these studies show that an increase in flow velocity or shear has been shown to increase the bacterial numbers in biofilms. Percival *et al.* (1999) and Ollos *et al.* (2003) indicated that this increase in biofilm biomass is especially high when the level of biodegradable organic matter is low. It is believed that biofilms rely on defensive mechanisms to resist detachment (Lehtola *et al.,* 2006). However, there are conflicting studies demonstrating that an increase flow velocity causes a reduction in the levels of biofilm formation (Donlan and Pipes, 1988). Alternatively, Tsai, (2006) showed that flow velocity had no significant effect on the formation of biofilms. These contradicting results may be attributed to the vast differences in various distribution systems and variations in the biofilm age. Ollos *et al.* (2003) acknowledges the fact that their results were based on young biofilms when compared to those existing in full-scale distribution systems. In addition, fluctuations in flow velocities results in resuspension of sediments and detachment of biofilm, both contributing to increases in turbidity and release of metals and organic matter (Batté *et al*., 2003; Lehtola *et al.,* 2006; Lui *et al.,* 2013a and 2014).

2.4.6 Residence Time and Water Age

Distances from treatment to tap has a significant effect on the residence time. The residence time of water within some distribution systems can reach up to a few days, leading to increased opportunities for microbial growth. In addition, the variability in pipe diameters and fluctuations in flow velocity caused by water consumption influences residence time specifically within reservoirs (Prest *et al*., 2016). Here, an increase in residence time correlates to an increase in bacterial abundances resulting in decay of disinfectant residuals. Wang *et al*., (2014) observed that changes in the water chemistry associated with increased water age such as decreases in disinfectant residual and dissolvable oxygen and an increase in TOC caused significant shifts in the microbial community.

2.4.7 Disinfectant Type – Chlorination and Chloramination

Of all the factors discussed above, disinfection of the drinking water can be considered the most significant in terms of controlling microbial growth and maintaining water quality within DWDS. Chlorine, chloramine and other chlorine compounds have long been successfully used for the control of microbial growth within DWDSs. Hwang *et al*. (2012b) observed that chlorination and chloramination treatments exerted strong selection pressures on the microbial community. Chlorine is an oxidising agent and it is known to be effective in injuring bacteria and thereby preventing and limiting their growth in the system (Vasconcelos *et al*., 1997). Chlorine exists in water as hypochlorite or hypochlorous acid which oxidises superficial biomolecules (membrane lipids and envelopes) as well as biomolecules within bacterial cells (enzymes and nucleic acids) resulting in cell death (Junli *et al*., 1997). For Chlorine to be effective, it must be present at a sufficient quantity and for a certain reaction time (Hwang *et al*., 2012b).

However, a chlorine residual (1 mg Cl_2/L) is often insufficient in killing and removing attached biofilms and a fraction of the bacterial population escapes disinfection (Norton and LeChevallier, 1997; Batté *et al*., 2003). Disinfectant residuals are typically maintained at levels lower than the maximum guidelines (chlorine: 5.0 mg/L

and chloramine: 3.0 mg/L) (WHO, 2011). The efficacy of disinfectants are influenced by biofilm formation and penetration of the disinfectant into biofilms as well as the presence of AOC and possible corrosion of the pipe surfaces (Norton and LeChevallier, 1997). It is important that a chlorine residual be maintained throughout the system yet there are multiple factors affecting the depletion of chlorine in the distribution system such as, composition of pipe materials, biofilms, presence of organic matter in the bulk water, size of the system, hydraulic forces and residence times (Norton and LeChevallier, 1997; Chu *et al.,* 2005). In addition, disinfection can potentially form carcinogenic disinfection by-products (DBP) such as trihalomethanes (THM), haloacetic acids (HAA), halonitromethanes (HNM), haloacetonitriles (HAN), etc. (Wilczak *et al.,* 1996; Goslan *et al.,* 2009; Bougeard *et al.,* 2010).

Although chlorination is successful in reducing bacterial growth, many distribution systems now apply chloramination as a second disinfection strategy (Neden *et al.,* 1992; Pintar and Slawson, 2003). Following and initial disinfection event, secondary disinfection aims to reduce microbial growth within the distribution system. In both cases disinfection greatly alters the composition and structure of the microbial community (Prest *et al*., 2016). Chloramines are typically applied in a 3:1 to 4:1 ratio of chlorine to ammonia. The ammonia associated with these ratios optimises the formation of monochloramines. An effective disinfectant should maintain its lethality by maintaining residual concentrations down the system. Chloramines have multiple benefits at the same concentrations of free chlorine. Chloramines are more stable than chlorine in maintaining residual disinfectant throughout the DWDS. They decrease heterotrophic plate counts (HPC) and coliform growth as well as improve the taste and odour of drinking water (Neden *et al.,* 1992). Volk and LeChevallier (1999) showed that the consumption of chloramine was lower than that of chlorine. They are not consumed by the biofilms extracellular polysaccharide matrix, they are able to penetrate the biofilm and interact specifically with the DNA and certain amino acids such as tryptophan and those containing sulphur (LeChevallier *et al.,* 1990; Geldreich, 1996). Chloramines can themselves produce THM and HAA, however they are only produced in trace amounts and much lower than that of chlorine (Norton and LeChevallier, 1997; Goslan *et al.,* 2009; Bougeard *et al.,* 2010).

2.5 CHALLENGES EXPERIENCED IN CONTROLLING MICROBIAL GROWTH

DWDS are complex aquatic environments and problems experienced cannot be dealt with in isolation. Many of the challenges experienced within DWDSs are microbially based (Lui *et al*., 2014), with some of the most common challenges associated with biofilm formation and microbial growth. The presence of biofilms and loose deposits act as sites for biomass accumulation and microbial mediated corrosion (Lui *et al*., 2013a). Furthermore, these sites of high biomass may harbour potential pathogens, increase depletion of disinfectant residual and deteriorate taste and odours.

2.5.1 Presence of Pathogens

The presence of biofilms offer a favourable environment and multiple advantages to the associated bacteria (Chao *et al*., 2015). They provide protection from environmental stresses and disinfection where nutrients and metabolic products are commonly shared and opportunities for horizontal gene transfer are improved with the drinking water community (Wingender and Flemming, 2011). As a result of these benefits, potential pathogens may reside in the biofilm where they are protected from residual disinfectant and are able to proliferate to higher abundances in the oligotrophic environment of the DWDS resulting in an increased public health risk (Chao *et al*., 2015). The growth and persistence of pathogens is a principle concern in DWDSs.

Pathogens such as *Legionella pneumophila, Pseudomonas aeruginosa, Aeromonas hydrophila, Klebsiella pneumoniae* and *Mycobacteria* spp. all have the ability to grow in the low nutrient environments of DWDS (Flemming *et al*., 2002; Wang *et al*., 2014; Prest *et al*., 2016). As well as bacterial pathogens, there are some protozoan pathogens such as *Cryptosporidium* and *Giardia*, which not only survive in drinking water systems but also potentially act as hosts for pathogenic bacteria such as *Legionella pneumophila* (Thomas and Ashbolt, 2011; Wang *et al.,* 2014). Lastly, the presence of viruses such as norovirus, hepatitis and rotavirus in drinking water has also been reported (Gall *et al*., 2015). Waterborne viruses have a range of different capsid protein structures and genome type and are able to persist in water for long periods of time (Gall *et al*., 2015).

2.5.2 Microbial Mediated Processes

Microbial mediated Corrosion

Beech and Sunner, (2004) defined biocorrosion or microbial influenced corrosion (MIC) as the accelerated deterioration of metals due to the presence of microorganisms on their surfaces. Biocorrosion is a result of the interactions between bacterial cells and their metabolites, abiotic corrosion products and the metal surface (Beech and Sunner, 2004). The development of biofilms on metal surfaces of drinking water pipelines can considerably alter the chemistry and kinetics of corrosion reactions on metal pipe surface leading to the potential acceleration or inhibition of corrosion (Sun *et al*., 2014). Problems associated with microbially mediated corrosion are more significant in iron and steel pipelines (Camper *et al.,* 2014). As mentioned previously, distribution systems with iron pipelines showed significantly higher biofilm densities compared to other materials (Niquette *et al.,* 2000). In addition, Sun *et al*. (2014) observed that the abundance of corrosion associated bacteria were significantly higher in biofilms originating from surface water than groundwater. Typically, the main types of bacteria involved in MIC in aquatic habitats include iron-oxidising / reducing bacteria, sulphate-reducing bacteria, sulphur-oxidising bacteria and manganese-oxidising bacteria (Beech and Sunner, 2004; Sun *et al*., 2014).

Bacteria may accumulate in tubercles created by corrosion of the iron pipe material (LeChevallier *et al*., 1987). Here, bacterial cells may be protected from environmental stresses and disinfection. Emde *et al.* (1992) observed a higher variety of species in corrosion-induced-deposits and in the bulk water phase following extended periods of chlorination. They concluded that corrosion tubercles are able to sustain a diverse

population of microorganisms including direct and opportunistic pathogens as well as creating a habitat for microorganisms that influence the water's taste and odour (e.g. *Actinomycetes* and fungi).

Nitrification

Chloramines are often added as secondary disinfectant when free chlorine residuals are difficult to maintain. Chloramines are produced from reactions with free chlorine and ammonia with monochloramine (NH_2Cl) being the most commonly used in drinking water treatment. Monochloramine has the same oxidising potential as free chlorine but is known to be more effective in reducing biofilm growth (Vikesland *et al*., 2001). Chloramines are considered to be more persistent that chlorine in drinking water (Norton and LeChevallier, 1997; Zhang *et al*., 2009). However, chloramination commonly causes undesirable nitrification, resulting in operational problems for many drinking water utilities (Regan *et al*., 2002). Nitrification then contributes to the depletion of monochloramine (Berry *et al*., 2006).

The introduction of ammonia into the system provides a potential source of nitrogen either by excess ammonia or through chloramine decay (Zhang *et al*., 2009). This promotes the growth of nitrifying microorganisms composing of different species of both bacteria and archaea (Belser, 1976; Nicol and Schleper, 2006). Bacterial nitrification in the DWDS causes an increase in nitrite and nitrate levels impacting water quality and negatively impacting infrastructure such as corrosion (Zhang *et al*., 2009; Wang *et al.,* 2014). In addition, nitrification results in the consumption of dissolvable oxygen and a decrease in pH (Kirmeyer *et al.,* 1995; Zhang *et al*., 2009). The loss of chloramine residual leads to an increase in heterotrophic bacterial concentrations and biofilm accumulation resulting in an increased potential for re-growth events within the distribution system (Kirmeyer *et al.,* 1995; Norton and LeChevallier, 1997; Pintar and Slawson, 2003).

Microbial nitrification is a two-step process: firstly ammonia (NH_4^+) is oxidised to nitrite (NO_2^-) by ammonia oxidising bacteria and archaea (AOB and AOA respectively) (Van der Wielen *et al*., 2009). Secondly nitrate is further oxidised to nitrite (NO_3^-) by nitrite oxidising bacteria (NOB) (Wolfe *et al.,* 1990; Cunliffe, 1991; Francis *et al.*, 2005). Nitrite in the system is problematic as it rapidly decreases the free chlorine and is also further oxidised leading to an accelerated decrease in residual chloramine (Wolfe *et al.,* 1990; Cunliffe, 1991). The free ammonia, nitrites and nitrates then serve as an energy source for AOB and NOB (Kirmeyer *et al.,* 1995; Pintar and Slawson, 2003).

Nitrifying bacteria are chemolithothrophic bacteria. Common bacterial genera involved in ammonia oxidation typically are *Nitrosomonas and Nitrosospira* and are members of the *betaproteobacteria* (Kowalchuk and Stephen, 2001; Zhang *et al*., 2009). In addition, multiple studies have identified AOA including members from the phylum *Thaumarchaeota* (You *et al*., 2009; Spang *et al*., 2010; Stahl and de la Torre, 2012). Nitrite oxidising bacteria associated with aquatic environments include members from the genus *Nitrobacter* belonging to the *alphaproteobacteria* as well as members from the genus *Nitrospira* (Zhang *et al*., 2009). However, recent studies have revealed the presence of the complete ammonia-oxidising (comammox) *Nitrospira*-like bacteria in DWDSs, which is capable of the completing the full oxidation of ammonia to nitrate

(Pinto *et al*., 2016). In light of this discovery, it is possible that comammox bacteria play a more significant role in nitrification that previously thought (Bautista-de los Santos *et al*., 2016).

Disinfection has the greatest influence on microbial growth within DWDSs and the microbial composition in the system is greatly altered with the change in disinfection strategy. It is therefore important to understand how disinfection affects the microbial community dynamics. However, it is apparent that of all these factors described above (pipe material, nutrient levels, temperature, hydraulic forces, disinfectant type, nitrification and microbial mediated corrosion), no one factor acts alone. These factors not only affect microbial growth but also influence the composition of the microbial community. The extent in which one factor influences the microbial composition, is itself affected (positively or negatively) by another and together they will have effect on the microbial growth within a DWDS. The DWDS is a multi-dimensional system where all contributing factors will have an overall combined effect on the microbial community and ultimately water quality. The complexity of interactions between the contributing factors, the great variability between different distribution systems and the difficulty in standardisation of those systems may explain the various conflicting results observed in literature. Reasons for the variable results found throughout can be attributed to the differences in source water quality and the various physio-chemical parameters that affect microbial growth and presence (Srinivasan *et al.,* 2008). It is therefore important to understand the impact of these multiple factors on the microbial ecology within DWDSs.

Increased Disinfectant Decay

Microbial growth influences the effectiveness of disinfectants as the effect of chlorine on penetrating biofilms is limited (Batté *et al*., 2003). Water suppliers aim to maintain a residual concentration of disinfectant to minimise the potential for microbial growth. However, with the concentrations of disinfectant used in DWDSs, biofilm development and microbial growth cannot be avoided (Rossman *et al*., 1994; Prest *et al*., 2016). Chlorine is a highly reactive chemical and readily reacts with a variety of inorganic and organic compounds thereby causing its gradual decrease in the distribution system (Vasconcelos *et al*., 1997). The consumption of chlorine in DWDSs is influenced by various factors namely through reactions with inorganic and organic chemicals, reactions with biofilms lining pipe walls, through the transport of chlorine in the bulk water and through reactions with corrosion process (Vasconcelos *et al*., 1997).

2.6 UNDERSTANDING MICROBIAL ECOLOGY IN DRINKING WATER DISTRIBUTION SYSTEMS

Drinking water is recognised as unique microbial ecosystem. As bacterial numbers are never eliminated from DWDS and due to the potential problems associated with microbial growth in the DWDS it is essential to fully understand the diversity and dynamics of the microbial community within drinking water treatment and distribution. Furthermore, drinking water utilities need to understand the specific dynamics inherent to each DWDS, how they drive changes in the microbial community and influence the microbial ecology.

2.6.1 Direct measurements and enumeration of bacterial concentrations

Traditionally heterotrophic plate counts (HPC) have been used to determine the efficacy of treatment in reducing bacterial concentrations and downstream monitoring bacterial levels in the DWDS. This culture-based test relies on the fact that heterotrophic bacterial, yeast and moulds require organic carbon for growth. However, only a small portion of metabolically active microorganisms in the water sample may grow. This leads to a great underestimation of the total microbial community present within the DWDS (Lui *et al.,* 2013a; Chiao *et al.,* 2014). HPC test are still employed today as indicators for the effectiveness of the water treatment process and therefore indirect indicators of pathogen removal and water safety.

To overcome the disadvantages of HPC, total cell counts (TCC) can be used. This method employs membrane filtration, fluorescent dye staining and microscopic counting to determine bacterial cell numbers (Boe-Hansen *et al.,* 2002). Fluorescent staining involves DAPI (4', 6-diamidino-2-phenylindole) staining which binds to double-stranded DNA and can pass through intact cell membranes giving an indication of the proportion of live cells in the samples. Stained cells can then be visualised microscopically.

More recently, flow cytometry (FCM) has great potential for total cell counts in drinking water (Berney *et al.,* 2008; Hammes *et al*., 2008). FCM is simple and rapid as well as more sensitive and accurate. Fluorescent labelling of nucleic acids allows for direct enumeration of total cell concentrations as well as detection of specific cellular features such as cell viability (Hammes *et al*., 2008). Typically, with the use of two stains, SYBR Green and propidium iodide (PI) cells can be differentiated as intact or damaged respectively.

Alternatively, the level of biologically active bacteria within the sample can be measured using adenosine triphosphate (ATP) as high ATP numbers correlate to high bacterial numbers (Lui *et al.,* 2013b). In addition, concentrations of AOC within the sample are used to determine the potential for microbial growth (Momba *et al.,* 2000; Ndiongue *et al.,* 2005). Hammes *et al*. (2008) compared data obtained from FCM with conventional HPC as well as ATP concentrations from different drinking water treatment processes. They observed that FCM showed clear advantages over HPC and ATP as it detects cells irrespective of culturability and that ATP measurements are often affected by extracellular ATP. In conclusion they suggest that total cell enumeration through FCM is a valuable tool in monitoring water quality treatment and distribution.

2.6.2 Techniques used in ecological studies of microbial communities: culture-independent techniques

Of the methods discussed above, none but HPC provides an indication of the microbial species present, their abundance or contribution to the total community within the distribution system. These methods may simply give an indication of cell numbers and cell viability, however they are unable to differentiate between bulk water bacteria or biofilm bacteria for example. As mentioned above that HPC is used for the detection of heterotrophic bacteria in DWDS, the resulting numbers are a great underestimation of the microbial community present as only the culturable microorganisms are identified. For this reason, culture-independent

and molecular methods have been developed. Using these methods, the microbial species present in the system are identified, specifically potential pathogens and their contribution to the microbiome evaluated.

Culture-independent methods involve direct DNA extraction directly from the sample and include methods such as denaturing gradient gel electrophoresis (DGGE), DNA fingerprinting, cloning and 16S rRNA gene directed PCR methods (Hoefel *et al.,* 2005). These methods concentrate on detection and identification of microbial species within DWDS. Commonly, T-RFLP analysis has been used to examine DWDS microbial communities (Hwang *et al.,* 2012a; Hwang *et al.,* 2012b; Wang *et al.,* 2012; Zhang *et al.,* 2012). Another fingerprinting method used is 16S rRNA single-strand conformation polymorphism (SSCP) based on extracted DNA and RNA which reveals bacterial species with relative high abundances (Schmeisser *et al.,* 2003; Eichler *et al.,* 2006; Henne *at al.,* 2012). Molecular fingerprints allow for evaluation of structural features such as relative abundance of a species, species richness and community composition (Henne *at al.,* 2012). Using fingerprints of both DNA and RNA, the species present are identified and the active members are assessed (Eichler *et al.,* 2006). However, these molecular methods could not represent the complete picture of diversity and community composition due to limited throughput (Sun *et al.,* 2014).

The use of the 16S rRNA gene in PCR-based fingerprinting techniques (i.e. DGGE, SSCP and T-RFLP) have been combined with molecular cloning and sequencing (i.e. conventional Sanger-sequencing) for the identification of bacterial species present in environmental samples. Although this approach has provided novel insight into the composition and structure of bacterial communities, it still underestimates the true diversity as it allows for the preferential selection and identification of dominant species only. In particular, Muyzer *et al*. (1993) reported that DGGE analysis only detected species that were present at a relative abundance of more than 1%.

2.6.3 Next generation sequencing driving the understanding of microbial communities and microbial ecology.

16S rRNA profiling

As early culture-dependent methods are limited in their ability to fully capture the complete microbial diversity within and environmental sample, it is therefore crucial to understand the bacterial population as a whole within an ecosystem, for example, which bacterial species are present in DWDS and how treatment and distribution system parameters effect their relative abundances and shape the bacterial community structure (Pinto *et al.* 2012). Recent developments in high-throughput and deep DNA sequencing (next generation sequencing, NGS) such as 454 pyrosequencing, Ion Torrent and Illumina MiSeq has greatly advanced our understanding of drinking water microbial ecology (Bautista-de los Santos *et al*., 2016). Next generation sequencing (NGS) targeting the 16S rRNA gene has highlighted the influence of environmental conditions (Pinto *et al*., 2014), disinfectant type (Gomez-Alvares *et al*., 2012; Hwang *et al*., 2012b), process operations (Pinto *et al*., 2012; Lautenschlager *et al*., 2014; Wu *et al*., 2015), hydraulic conditions (Douterelo *et al*., 2014), pipe material and distribution system infrastructure (Yu *et al*., 2012; Wang *et al*., 2014), etc. on the bacterial community structure.

The major advantages of NGS platforms are that (i) it avoids the need for laboratory isolation and cultivation, as it allows for the direct amplification of nearly all 16S rRNA genes from environmental samples, using the total DNA extracts obtained; (ii) there is no need for vector-base cloning, as it allows for *in vitro* separation of the 16S rRNA amplicons; and (iii) it allows for the sequencing of multiple samples in parallel, which is faster and more cost effective (Glenn, 2011; Liu *et al*., 2012b; Shokralla *et al*., 2012). One disadvantage, however, is that the current NGS platforms can only sequencing a short region of the 16S rRNA gene, which often does not provide sufficient taxonomic resolution beyond the level of family (Foster *et al*., 2012).

Initial bacterial community profiling studies used 454 NGS platforms (i.e. GS-FLX, GS-FLX+ etc.) (Roche Diagnostics Inc., Basel, Switzerland) that generated about 1 million reads per run, with read lengths that ranged between 600 bp and 800 bp (averaged, 500 bp). However, there has been growing interest in using other NGS platforms, in particular, the Illumina MiSeq NGS platform (Illumina 7 million reads per run, with read lengths ranging between 100 bp and 250 bp) (Glenn, 2011; Liu *et al*., 2012b; Shokralla *et al*., 2012). These NGS platforms, although different, share three main steps: (i) library preparation and clonal amplification, (ii) sequencing and (iii) base-calling and visualisation (Mardis, 2008; Glenn, 2011; Liu *et al*., 2012b).

Originally used 454 NGS platforms (i.e. GS-FLX, GS-FLX+ etc.) utilised emulsion PCR for library preparation and clonal amplification, and pyrosequencing chemistry for sequencing (Mardis, 2008; Metzker, 2010). Emulsion PCR occurs on the surface of beads that are compartmentalised in water droplets, also called micro-reactors, in a water-in-oil emulsion (Berka *et al*., 2010; Kanagal-Shamanna, 2016). The surfaces of the beads are decorated with numerous, randomly scattered oligonucleotides that are complementary to specific adaptor sequences, which are either ligated or PCR amplified to the ends of the amplicons. This complementarity ensured hybridisation of the amplicons to the surface of the beads, and subsequent amplification to generate multiple copies of the same amplicon per bead (Metzker, 2005; Fedurco *et al*., 2006; Mardis, 2008).

Multiple studies have made use of 454 pyrosequencing for microbial community and diversity analyses of both bulk water and biofilms in DWDS (Hong *et al.,* 2010; Kwon *et al.,* 2011; Hwang *et al.,* 2012b; Gomez-Alvarez *et al.,* 2012; Pinto *et al.,* 2012; Zhang *et al.,* 2012; Chiao *et al.,* 2014; Douterelo *et al.,* 2014; Sun *et al.,* 2014). All these studies demonstrate the ability of pyrosequencing to characterise microbial communities within DWDS revealing how the disinfectant regimes and various environmental factors influence and change the community composition and diversity. Here, taxonomic identification is improved and rare species are identified. Pinto *et al*. (2014) used 454 pyrosequencing to show the temporal and spatial dynamics of bacterial communities within a drinking water system. With the resulting 16S rRNA profiles they were able to show that the bacterial community demonstrated trends based on seasonal cycles as well as the type of source water.

More recently, the development of Illumina MiSeq technology enables high-resolution characterisation of microbial communities with read lengths comparable to 454 pyrosequencing at a lower cost (Caporaso *et al*.,

2012). Fadrosh *et al.* (2014) describes a dual-indexing amplification and sequencing approach to assess the composition of the microbial community using the Illumina MiSeq platform. Here they confirm that this approach provides a cost effective and flexible sequencing option. Illumina is now replacing 454 pyrosequencing as the method of choice for microbial community related studies (Douterelo *et al*., 2104). Illumina utilises bridge or solid phase amplification for library preparation and clonal amplification, and cyclic reversible termination chemistry for sequencing (Mardis, 2008, Metzker, 2010). Bridge amplification occurs on the surface of a flow cell that is decorated with numerous, randomly scattered oligonucleotides. These oligonucleotides are complementary to specific adaptor sequences that are either ligated or PCR amplified to both ends of the amplicons. Following hybridisation, bridge amplification commences and multiple copies of the same amplicon are generated per cluster (Metzker, 2005; Fedurco *et al*., 2006; Mardis, 2008).

Here, community profiling studies using Illumina MiSeq, have revealed shifts in the microbial ecology due to sampling location, treatment and processing as well as temporal variations (Baron *et al*., 2014; Roeselers *et al*., 2015; Wu *et al.* 2015). Kozich *et al*. (2013) presented a method for sequencing hypervariable regions within the 16S rRNA gene using the Illumina MiSeq NGS platform. This method has been used in numerous studies of various environmental settings to characterise bacterial communities, and to explain compositional and structural changes across different spatial and temporal scales (Barret *et al*., 2015; Fuhrman *et al*., 2015; Hong *et al*., 2015; Liu *et al*., 2015). This method, in particular, allows for the generation of overlapping, paired 250 bp reads of hypervariable regions (i.e. V34, V4, and V5) within the 16S rRNA gene, producing high quality sequence data that are comparable, and often better that those of 454 NGS platforms (Luo *et al*., 2012; Kozich *et al*., 2013). Moreover, this method provides higher sequencing coverage per sequence run and per cost, when compared to 454 NGS platforms, which allows for more detailed and accurate descriptions of bacterial communities (Glenn, 2011; Liu *et al*., 2012b; Shokralla *et al*., 2012; Kozich *et al*., 2013).

Metagenomic approaches

Shotgun metagenomic sequencing is a relatively new and robust approach to environmental sequencing, shedding light on microbial community biodiversity and function. This approach has uncovered the extensive biodiversity within microbial communities as well as microbial ecology in terms of the functions behind the interactions between individuals and their environment (Sharpton, 2014). Here DNA is extracted directly from the environment, from all cells in the community. Total genomic DNA is then sheared into smaller fragments that are sequenced individually. The resulting reads consist of DNA sequences from various regions of the genomes of different individuals within the community. Some reads will be taxonomically informative where as others will provide insight into the biological functions encoded in the genomes of the community members (Sharpton, 2014). Here shotgun metagenomic sequencing provides the opportunity to answer two questions, i.e. who are the community members and what are their functions?

Metagenomes provide insight into the collective functions encoded in the genomes of community members. Metagenomic reads containing protein coding sequences are identified and the potential gene function is predicted by comparing sequences to databases of genes, proteins or protein families and metabolic pathways. This potentially produces a functional profile of the community, which can then be compared to

other communities, revealing functions that are associated with a specific environment (Sharpton, 2014). Furthermore, genomic binning of metagenomic data can be applied to investigate individual genomes of unculturable microorganisms from various ecosystems and reconstruct partially-complete genomes of dominant members of the community (Allen and Banfield, 2005; Chao *et al*., 2015). Using Illumina metagenomic data, Chao *et al*., (2015) were able to identify some of the key metabolic functions of the drinking water biofilms as well as reconstruct partial genome of *Bradyrhizobiaceae*-like bacterium. From shotgun DNA sequencing of biologically active filters in a drinking water treatment plant, Pinto *et al*. (2015) were also able to discover a metagenomic bin closely relating to the comammox *Nitrospira*-like bacteria, capable of complete oxidation of ammonia to nitrate.

The community profile data generated from high throughput sequencing provides insight into the biodiversity within and between microbial communities. Microbial biodiversity is not evenly distributed as different sites typically contain different microbial communities. Therefore, measuring the differences between communities can improve the understanding of how the biodiversity is distributed. Researchers are able to observe the species present and their abundance in the community at a specific site (*alpha*-diversity) as well as compare the species diversity between different sites or the degree of community differentiation (*beta*-diversity) (Tuomisto, 2010).

2.7 CONCLUSION

Drinking water treatment utilities aim to produce water that is operationally and aesthetically acceptable and free from pathogens, from treatment, through the DWDS and to the consumers tap. Based on information currently available, consecutive treatment operations employed in DWTPs not only play an essential role in improving the quality of water, but also influence the microbial community that migrates from the DWTP into the DWDS. Previous community composition studies of bacterial communities have attempted to answer questions such as ‘what species are present in the community?’ and ‘how does the community change across different spatial and temporal scales?’ However, few of these studies have attempted to assess the influence of different physicochemical water characteristics, alone or in combination, on the bacterial community. The opportunity to gain novel insight into the microbial ecology of drinking water is often lost when the physicochemical factors affecting the microbial community are not considered.

Presently, little is known about the spatial heterogeneity and temporal variability bacterial communities associated with treatment of source water, specifically RGS filters as well as the DWDS. This is particularly relevant, as changes in the RGS filter bacterial community may have an effect on essential microbial-mediated processes occurring during filtration, play a role in improving the quality of water, and also influence the composition and structure of downstream bacterial communities in the DWDS. The majority of studies performed on South African water systems have concentrated on water in rural communities, the presence of waterborne pathogens and their associated health risks. Information on the microbial ecology of developed South African drinking water distribution systems is therefore sorely lacking with limited understanding in

microbial profiling in South African drinking water treatment and distribution. Understanding the microbial community dynamics in associated with these water environments can greatly improve water management. Setting up a predictive framework will help in eliminating microbial risks as well as upgrade water quality monitoring methods, making them more resource efficient and sustainable.

CHAPTER 3: SPATIAL AND TEMPORAL STRUCTURE OF BACTERIAL COMMUNITIES HARBORED WITHIN RAPID GRAVITY SAND FILTERS

3.1 INTRODUCTION

Rapid gravity sand (RGS) filtration is a physical treatment operation step employed in DWTPs, whereby the source water, obtained from surface water systems (i.e. rivers, dams and lakes) and/or groundwater systems, is passed through a porous granular medium, often consisting of sand (Stevenson, 1994; Davies and Wheatley, 2012). RGS filtration systems, including rapid sand (RS) filters and granular activated carbon (GAC) filters, are designed to facilitate the removal of natural (or total) organic matter constituents, including dissolved organic matter (i.e. dissolved organic carbon, nitrogen and phosphorous) and undissolved organic matter (i.e. particulate organic carbon) (Pagano *et al*., 2014). The purpose of sand filtration is to deliver low-nutrient oligotrophic filtered water that reduces the growth potential of microbial species during drinking water distribution (Liu *et al.*, 2012; Bai *et al*., 2013; Feng *et al.*, 2013; Liao *et al*., 2013; Lautenschlager *et al.*, 2014; Liao *et al*., 2015b).

In addition to the removal of organic constituents, RGS filtration systems facilitate the removal of microbial species. Some species attach to the RGS filter particles and consequently become spatially stratified along the vertical depth of the RGS filter bed with increased operating time (Liu *et al*., 2012; Bai *et al.*, 2013; Gülay *et al.*, 2016). These microbial species often persist on the filter bed in complex, metabolically active microbial communities that maintain cell concentrations ranging from 10^{15} to 10^{16} cells/m^3 (Magic-Knezev and Van Der Kooij, 2004; Velten *et al*., 2007; Bar-Zeev *et al*., 2012; Lautenschlager *et al*., 2014). Furthermore, these microbial species play an important role in improving the quality of water, as they readily degrade the organic matter retained in the RGS filter bed as part of their metabolic activity. Many attempts have been made to identify the metabolic capabilities of RGS filter microbial species (Albers *et al*., 2015; Bai *et al*., 2013; Tekerlekopoulou *et al*., 2013; Gülay *et al*., 2016). For example, several studies have identified *Nitrospira,* of the phylum *Nitrospirae,* as the main nitrite-oxidising bacteria in RGS filter systems (De Vet *et al.*, 2009; White *et al.*, 2012; Albers *et al*., 2015; LaPara *et al.*, 2015; Gülay *et al.*, 2016), while *Nitrosomonas,* of the phylum *Proteobacteria,* are identified as the main ammonia-oxidising bacteria in RGS filter systems (De Vet *et al*., 2009; Albers *et al*., 2015; Gülay *et al*., 2016). Furthermore, within RGS filter systems, species of the genus *Gallionella* are seemingly involved in iron oxidation, whereas species of the genera *Hyphomicrobium* and *Pedomicrobium* are involved in manganese oxidation, and methanotrophic species of the family *Methylococcaceae* are involved in methane oxidation (De Vet *et al*., 2009; Albers *et al*., 2015, Gülay *et al*., 2016).

Moreover, RSG filter systems influence the DWDS microbial community, as it maintains a stable microbial community from which selected RGS filter-independent microbial species are constantly seeded (Pinto *et al.*, 2012; Zeng *et al.*, 2013; Lautenschlager *et al.*, 2014; Gülay *et al.*, 2016; Li *et al.*, 2017). Although stable, RGS filter microbial communities are influenced by several factors, including differences in RGS filter design (Lautenschlager *et al.*, 2014), treatment operations prior to RGS filtration (Fonseca *et al.*, 2001; Lautenschlager *et al.*, 2014), physical and chemical water properties (Li *et al.*, 2010; Pinto *et al.*, 2012; Liao *et al.*, 2013, Kim *et al.*, 2014; Albers *et al.*, 2015) and periodic backwashing (Kasuga *et al.*, 2007; Liu *et al.*, 2012; Gibert *et al.*, 2013; Liao *et al.*, 2015b).

Recently, the characterisation of microbial species, particularly microbial species, in RGS filter systems has been done through conventional culture-independent profiling techniques (Fonseca *et al*., 2001; Kasuga *et al.*, 2007; Feng *et al.*, 2012; Liao *et al.*, 2012; Feng *et al.*, 2013) and next generation sequencing (NGS) technologies (i.e. 454-pyrosequencing, Illumina MiSeq and Ion Torrent) (Kwon *et al.*, 2011; Pinto *et al.*, 2012; Bai *et al.*, 2013; Kim *et al.*, 2014; Lautenschlager *et al.*, 2014; Lin *et al.*, 2014; Albers *et al.*, 2015; Gülay *et al.*, 2016). In general, NGS technologies are the preferred choice, as they enable high-throughput and high-resolution characterisation of microbial species without prior cultivation and separate DNA isolations (Foster *et al.*, 2012; Douterelo *et al.*, 2014; Proctor and Hammes, 2015). Studies employing NGS technologies have shown that DWTP RGS filters harbour complex microbial communities that comprise of up to 52 microbial phyla, and between 500 and 1 000 microbial species or operational taxonomic units (OTUs) defined at a 97% sequences similarity threshold (Kwon *et al.*, 2011; Pinto *et al.*, 2012; Kim *et al.*, 2014; Lautenschlager *et al.*, 2014; Lin *et al.*, 2014; Albers *et al.*, 2015; LaPara *et al.*, 2015; Gülay *et al.*, 2016). Although compositionally diverse, RGS filter systems of different DWTPs tend to be dominated by *Proteobacteria* and share numerous other phyla, including *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Firmicutes*, *Gemmatimonadetes*, *Nitrospirae*, *Planctomycetes*, and *Verrucomicrobia* (Kwon *et al.*, 2011; Pinto *et al.*, 2012; Bai *et al.*, 2013; Kim *et al.,* 2014; Lautenschlager *et al.*, 2014; Lin *et al.*, 2014; Albers *et al.*, 2015; Gülay *et al.*, 2016), all of which are well-known freshwater (i.e. rivers, lakes and dams) phyla (Newton *et al.*, 2011; Martinez-Garcia *et al*., 2012; Llirós *et al.*, 2014).

Based on the findings of Pinto *et al*. (2012), predicting the population in drinking water distribution systems is a possibility and could provide important and relevant information for the control and management of the microbial water quality in the system. This would, however only be possible if the sand filter community is homogenously distributed across the filter at a specific point in time. In this study, the microbial communities of rapid sand (RS) filters harboured within the different filter galleries of two large-scale DWTPs were characterised by high-throughput sequencing of the V4-hypervariable region of the 16S rRNA gene. The primary objectives of this study were to determine the extent of spatial variation in the microbial community structures and composition (i) along the surface and depth of an individual RS filter bed, (ii) across parallel RS filters within a filter gallery, and (iii) across filter galleries; and, moreover, to determine (iv) the extent of temporal variation in the microbial community structures.

3.2 MATERIALS AND METHODS

3.2.1 Study sites, sample collection

Drinking Water Treatment Plant 1

For the first treatment plant, sampling was conducted on a monthly basis from May 2015 to September 2015. Samples were collected at designated RS filters harboured within the filter galleries of purification stations 1, 3 and 4 (i.e. PS1, PS3 and PS4) (Figure 3.1). This treatment plant treats surface water, extracted from a large dam and combines six conventional treatment operations, i.e. coagulation, flocculation, sedimentation, carbonation, filtration (i.e. RS filtration) and disinfection (chlorination).

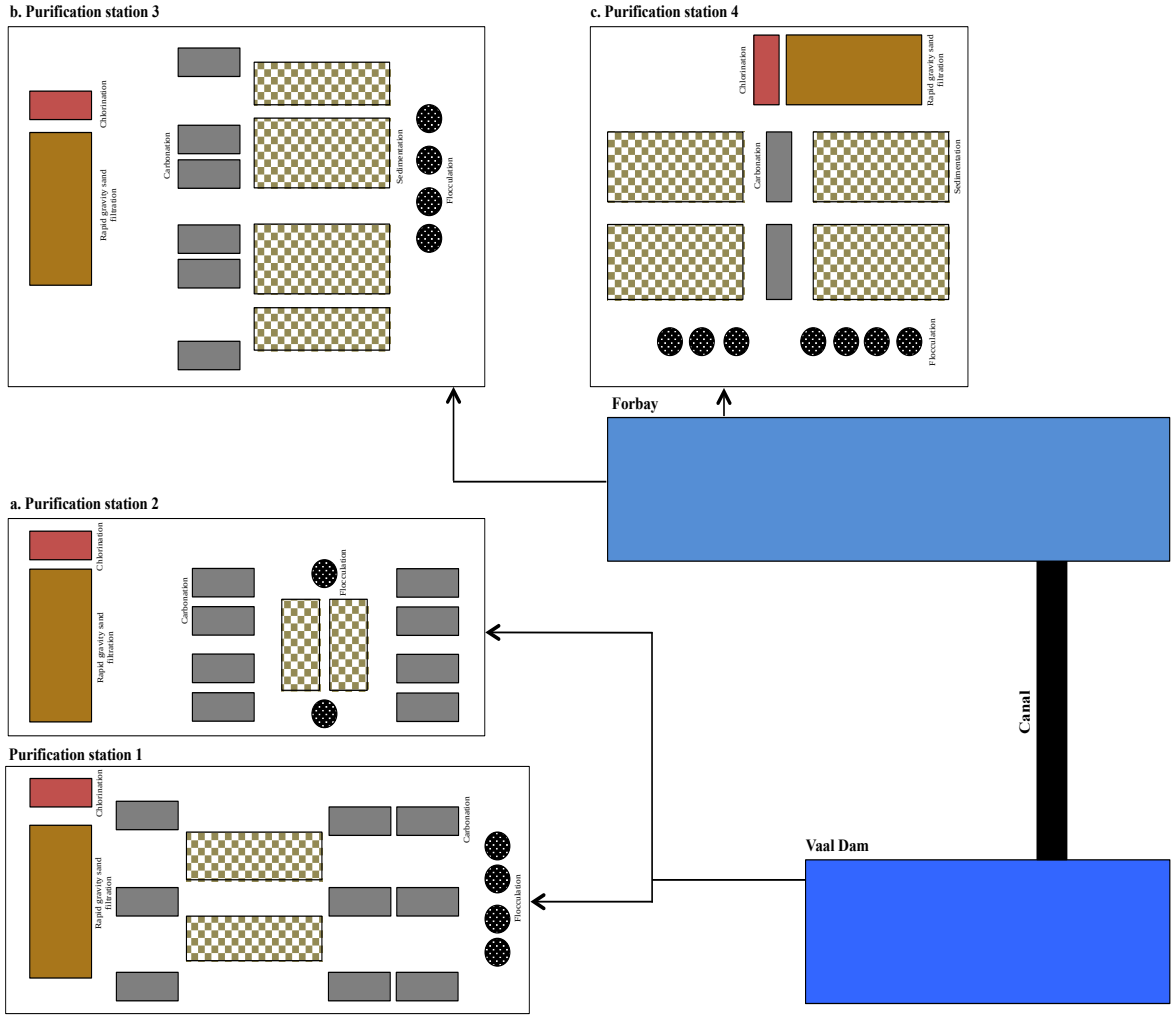


Figure 3.1 Schematic representation of the first drinking water treatment plant. Sampling was conducted over four months on a monthly basis at designated rapid sand filters harbored within the filter galleries of purification station 1, 3 and 4.

Within each purification station (PS), three RS filters (that were backwashed approximately 34 hours earlier) were randomly selected, drained onsite, and sampled according to a fixed sampling regime. The sampling regime was designed to study the extent of spatial variation in the microbial communities located: (i) along the surface and depth of the RS filter bed, (ii) across parallel RS filters within a filter gallery, and (iii) across different filter galleries. To study the extent of variation along the surface and depth of the RS filter bed, one RS filter within each PS was randomly selected and sampled extensively (i.e. RS filter 46 in PS1, RS filter 93 in PS3 and RS filter 66 in PS4). From these selected filter beds nine samples were collected along the surface of the filter beds as well as a core sample (estimated depth of 40 cm) from the centre as depicted in Figure 3.2.

Altogether, 130 filter bed (FB) surface samples ($n_{samples}$ = 43 for PS1, 44 for PS3 and 43 PS4,) and 12 core samples ($n_{samples}$ = 4 for PS1, 4 for PS2 and 4 PS4,) were collected (Table A1). To study extent of variation within a single filter gallery, and across filter galleries, FB samples were collected from two additional RS filters (RS filters 49 and 50 in PS1; RS filters 94 and 95 in PS3; and RS filters 51 and 54 in PS4) within each PS at a single location. However, two FB samples were not collected during the sampling period due to logistical difficulties (i.e. PS1_RSF50_S5_JUN_FB59 and PS1_RSF50_S5_JUL_FB107). Approximately 40 g of FB surface samples were collected and deposited in a sterile 1 litre Nalgene™ polycarbonate bottles (Thermo Scientific™, South Africa), using an ethanol sterilised metal spatula. All samples were immediately transported to the laboratory, stored at 4°C and processed within 48 hours.

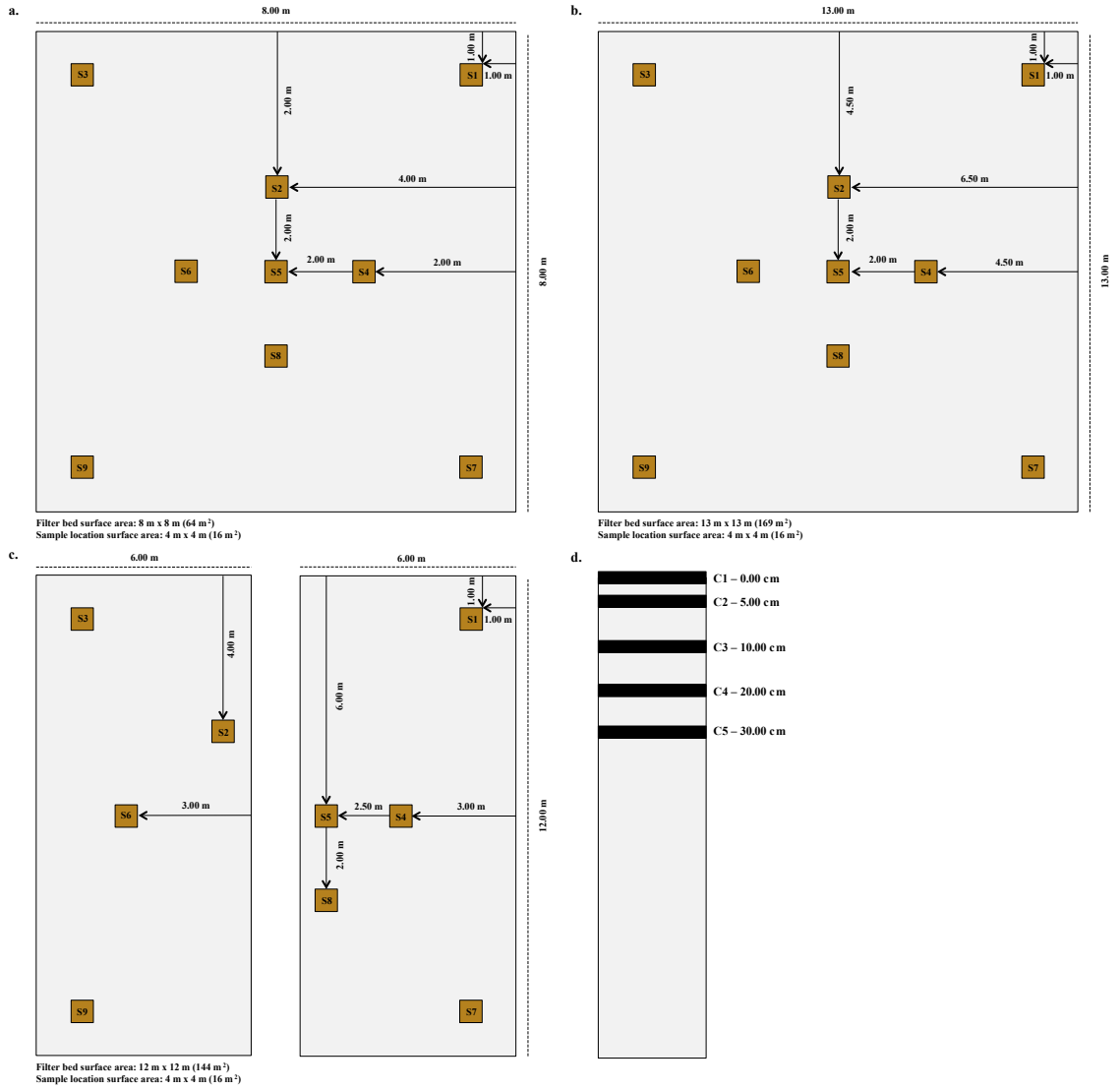


Figure 3.2 Schematic representation and specifications of the surface area of (a) RS filter 46 of PS1, (b) RS filter 93 of PS3, and (c) RS filter 66 of PS4; that were sampled monthly over 4 months along the surface and depth of the RS filter bed. The nine surface locations, designated S1 to S9 are indicated with red squares; and (d) the five core locations, designated C1 to C5 with associated depths are indicated with broad black lines.

Drinking Water Treatment Plant 2

The filtration process at the second DWTP is carried out at two purification systems or filter houses. However, in this study, FB sand samples were only collected from one of these filter houses, namely filter house 3. Within filter house 3, water is filtered by six galleries each consisting of 12 rapid sand filters (total of 72 filters), each with a filtration area of 148.78 m^2. The filter media of these RS filters consists of 5 layers, i.e. the top layer, which is called a support layer followed by a stone, coarse grit, fine grit and a coarse and fine sand layer. The incoming source water filters through the RS filter at a 4.0 m/h flow rate. The RS filters are covered, whereby light is excluded to less than 25 lux to prevent algal growth on the filter beds. The filters are cleaned every 24 hours by means of backwashing with chlorinated water. During backwashing the RS filters are aerated (air scour rate: 27 m/h), i.e. air is pump from the bottom of the filter upwards, for 4 minutes

and then washed with chlorinated water at a rate of 32 m/h to loosen all the collected dirt from the filter media.

Sampling was done on a monthly basis over a six-month period (February 2016 to July 2016). Samples were collected in a similar way as for the first treatment plant (DWTP 1) as depicted in Figure 3.2. Specifically, sand media was collected from three randomly designated RS filter following a specific sampling regime. This sampling regime was formulated in such a way to investigate the possible homogeneity across the surface of a single RS filter bed (FB_174), along the depth of the same RS filter as well as between three separate filter beds (FB_174, FB_179 and FB_180) within the same filter house. With respect to FB_174, RS filter samples were collected from nine selected areas on the filter bed surface and a core sampler was used to collect depth samples at the centre of the filter bed along different depths (e.g. 0 cm, 5 cm, 10 cm, 20 cm and 30 cm). With respect to FB_179 and FB_180 only one sample was collected in the middle of these RS filter beds (Figure 3.4). Approximately 40 g of surface sand samples were collected and placed into sterile 50 ml Falcon tubes using an ethanol sterilized metal spatula. The RS filter samples were transported to the laboratory where it was kept at 4 °C and processed within 24 hours. A total number of 94 FB media samples were collected (66 were surface samples 28 depth samples). During the months of April and June depth samples were collected only up until 20 cm thus no 30 cm depth samples were collected.

3.2.2 Sample preparation and processing

Filter bed media sample preparation and processing

To obtain microbial biomass from sand samples, 10 g of the filter media was mixed with 50 ml extraction buffer consisting of 0.4 g/L EGTA (Ethylene glycol-bis (2-aminoethylether)-N,N,N′,N′-tetraacetic acid), 1.2 g/L TRIS, 1 g/L peptone and 0.4 g/L N-dodecyl-N, N dimethyl-3-amminio-1-propanesulfonate. Thereafter, the samples were sonicated for 1 min to detach microbial cells potentially adhering to the sand particles (Camper *et al*., 1985). After sonication, the aqueous phase was manually filtered through a Sterivex™-GP 0.22 µm polycarbonate membrane filter unit (Merck Millipore, South Africa), using a sterile 20 ml syringe. The filter units were transferred to sterile glass screw cap bottles and stored at 4 °C until DNA extraction.

To obtain FB samples from the core samplers, the core FB media was removed and aseptically separated into five sections (0 cm, 5 cm, 10 cm, 20 cm, and 30 cm) (as indicated as indicated in Figure 3.2 d). The FB media from one core sampler that was collected in July from RS filter 46 located within PS1 was only separated into 4 sections (0 cm, 5 cm, 10 cm, and 20 cm). After separating the FB media from the core samplers, microbial biomass was obtained from all FB samples using the standardised procedure as described above.

Genomic DNA extraction

Genomic DNA was extracted using the phenol-chloroform DNA extraction protocol as described by Urakawa *et al*. (2010), with modifications incorporated by Feinstein *et al*. (2009) and Pinto *et al*. (2012)**.** Prior to DNA extraction, the polycarbonate filter with collected microbial biomass was cut into several pieces using a

sterile scalpel. The filter pieces were transferred to 2 ml Lysing Matrix Tube E tubes (MP Biomedical, South Africa), using a sterile tweezer and with the addition of 300 µl of 2 x TENS buffer (100 mM Tris-Cl, 40 mM EDTA, 200 mM NaCl, 2% SDS) and 900 µl phenol:chloroform:isoamyl alcohol (25:24:1, pH 8). The mixture was then subjected to three consecutive bead-beating steps, using the Tissue-Lyser II, at the highest frequency (setting 6) for 40 sec. After the initial bead-beat step, the homogenized mixture was centrifuged for 10 min at 14 000 x g using the Hermle Z 200™ centrifuge, and the aqueous phase was transferred to a 2 ml Eppendorf tube. The original 2 ml Lysing Matrix Tube E tubes, from where the aqueous phase was removed were then re-filled with 200 µl TENS buffer, prior to the initiation of the next bead-beating step. The bead-beating step was repeated two more times with centrifugation at 12 500 x g for 10 min. After bead-beating, the collected aqueous phase (approximately 600 µl) was supplemented with 900 µl phenol-chloroform-isoamyl alcohol (Sigma Aldrich, South Africa), mixed by repeated inversions and then centrifuged at 14 000 x g for 5 min. The resulting aqueous phase was transferred to a 2 ml Eppendorf tube and treated with ½ volume (350 µl) 7.5 M ammonium acetate (Sigma Aldrich, South Africa) and 600 µl chloroform (Sigma Aldrich, South Africa). The tubes were mixed by repeated inversions and centrifuge for 14 000 x g for 5 min. The aqueous phase was transferred to a 2 ml Eppendorf tube and incubated with ½ volume isopropanol (600 µl) (Sigma Aldrich, South Africa) and 6 µl GlycoBlue™ coprecipitant (15 mg/ml) (Thermo Fischer Scientific, South Africa) at -70°C for 10 min. The mixture was then centrifuged for 30 min at 12 000 x g at 4°C using the Eppendorf Centrifuge 5804R™, for nucleic acid precipitation. The supernatant was removed and the blue pellet was washed with 1 ml 80% ethanol and centrifuged for 30 min at 4 °C using the Eppendorf Centrifuge 5804 R™. The supernatant was removed, the pellet was air dried and subsequently re-suspended in 50 µl nuclease free water (Qiagen, South Africa) and stored at -20°C until further processing. DNA concentration and purity were determined using the Nanodrop ND-1000™ Spectrophotometer (Thermo Scientific, South Africa).

16S rRNA gene amplification and Illumina MiSeq sequencing

As a quality control step, to confirm that DNA extraction was successful and free of inhibitors, Polymerase chain reaction (PCR) was performed on the total genomic DNA extracts, using primers targeting the 16S small ribosomal ribonucleic acid (rRNA). The primers were 27F (5'-AGA GTT TGA TCC TGG CTC AG-3′) and 1492R (5'-GGT TAC CTT GTT ACG ACT-3') based on Edwards *et al*. (1989). PCR reactions were carried out using the BIO-RAD T100™ Thermal Cycler. The PCR mixtures (25 µl) consisted of 1 x reaction buffer, 1.50 mM $MgCl_2$, 250 µM of each nucleotide (dATP, dCTP, dGTP, dTTP), 10 pmol of each primer (forward and reverse), 1.50 U Taq DNA polymerase, 16.85 µl nuclease free water (Qiagen, South Africa) and 0.50 µl genomic DNA. The cycling conditions for the 16S rRNA amplicons consisted of an initial denaturation step at 92°C for 10 min, followed by 30 cycles of denaturation at 92°C for 1 min, annealing at 58°C for 1 min, extension at 75°C for 1 min, and a final extension at 75°C for 5 min. At the end of the 30 cycles, the reaction was kept at 4°C. When necessary, the PCR was repeated for samples that had an initial failed amplification, by diluting the DNA (1:10) with nuclease free water (Qiagen, South Africa) to remove possible inhibitors (i.e. humic substances, phenolic compounds). The genomic DNA of samples which had successful 16S rRNA amplification, were sent to the University of Michigan Medical School (Ann Arbor, United States of America)

for sequencing of the V4 hypervariable region of the 16S rRNA gene, using the Illumina MiSeq platform with a pair-ended sequencing protocol as described by Kozich *et al*. (2013).

3.2.3 Sequence processing and data analysis

MOTHUR sequence processing

Sequence processing and data analyses were conducted using the MOTHUR software package (version 1.35.1) and processing pipeline as described on the MOTHUR website (http://www.mothur.org/wiki/MiSeqSOP) (Schloss *et al*., 2009). Briefly, contigs were assembled between corresponding reverse and forward sequence pairs, and all sequences that were not assembled as well as those assembled with an insufficient overlap were removed (function: make.contigs). Furthermore, any sequences with ambiguous bases and sequences with a base pair length less than 275 were removed (function: screen.seqs). After initial quality filtering, the sequences were aligned to a reference alignment that was generated from the SILVA seed ribosomal RNA database (function: align.seqs) (Pruesse *et al*., 2007). All sequences that did not align were removed, and the aligned sequences were subsequently trimmed to ensure that the sequences all start and end at the same alignment coordinates (start = 11 894; end = 25 319) (function: screen.seqs). For computational purposes, duplicate sequences and sequences with a 2-bp similarity threshold were merged (function: unique.seqs and pre.cluster) (Huse *et al*., 2010). The resulting merged sequences were screened for chimeras using UCHIME (function: chimera.uchime) and removed (Edgar *et al*., 2011). Taxonomic affiliation was assigned to each of the chimeric-free sequences using the Greengenes reference taxonomy database with a pseudobootstrap confidence score of 80% (function: classify.seqs). Unwanted lineages were removed by eliminating sequences that could not be classified to kingdom level, or that classified as Eukaryota, chloroplasts or mitochondria (function: remove.linage). Following the removal of unwanted linages, the sequences were again classified using the Greengenes reference taxonomy database (function: classify.seqs). To obtain an operational taxonomic unit (OTU) table, the sequences were split into groups corresponding to their taxonomy at the level of order and assigned to OTUs using a 97% sequence similarity threshold (functions: cluster.split and make.shared). From these OTUs, representative OTU sequences were extracted and classified using the Greengenes reference taxonomy database (gg_13_8_99) (functions: get.oturep and classify.seqs).

Alpha and beta diversity analysis

Alpha and beta diversity indexes were calculated using functions provided in the MOTHUR software package (version 1.35.1) (Schloss *et al*., 2009). Three alpha diversity indexes including observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness index (J) were calculated using the summary.single function with incorporated parameters; iters = 1 000 and subsampling = T (sample containing the least amount of sequences). S_{obs} was used as a measure of within community species richness, whereas Shannon diversity index (H') and Pielou's evenness index (J) were used as measures of within community heterogeneity and evenness, respectively. In addition, Good's coverage scores were calculated to assess whether sufficient sequences were retained after subsampling for an accurate representation of the variation within each sample.

For beta diversity analyses, spatial and temporal variability in the compositional profiles of the communities were analysed using four ecological coefficients of compositional dissimilarity: OTU-based dissimilarity matrices Jaccard and Bray-Curtis, and phylogeny-based dissimilarity matrices weighted UniFrac and unweighted UniFrac (Bray and Curtis, 1957, Jaccard, 1912, Lozupone *et al*., 2011). Jaccard and unweighted UniFrac matrices were used in the analysis of community membership, as calculated pair-wise dissimilarity among selected samples is based on incidence data (presence and absence), whereas, Bray-Curtis and weighted UniFrac matrices, were used for the analyses of community structure, as pair-wise dissimilarity between selected samples is calculated on the basis of incidence and abundance data. The Bray-Curtis and Jaccard matrices were obtained using the dist.shared function with incorporated parameters, iters = 1 000 and subsampling = T (sample containing the least amount of sequences). Weighted and unweighted UniFrac matrices were obtained by constructing a phylogenetic tree with representative OTU sequences (output generated from get.oturep) using the dist.seqs and clearcut functions (Evans *et al*., 2006). Weighted and unweighted UniFrac matrices were then generated using functions unifrac.weighted and unifrac.unweighted.

In order to minimise the effect that rare species can have on the analysis of beta diversity the original data for the second treatment plant was truncated by removing OTUs which had an abundance of less than 1% of the total number of sequences in the dataset, by setting the threshold at 99% removal (Gobet *et al*., 2010). This dataset was further truncated by retaining only those OTUs, which had a frequency of 75% in all samples by using the MOTHUR filter.shared command with the minimumsample option. After generating this smaller dataset beta diversity analyses continued. A Mantel's test was performed in order to determine whether this sub-set of the data maintained the variation and was still representative of the whole RSF microbial community.

Statistical analysis

Multiple statistical tools were used in MOTHUR (version 1.35.1) (Schloss *et al*., 2009) and R (Venables *et al*., 2004). One-way analysis of variance (ANOVA) was performed in R, using functions lm and anova provided in the stats package, to compare the mean values between spatial groupings corresponding to samples obtained: (i) along the surface of the RS filter (S1-S9, as displayed in Figure 3.2 plots a, b and c), (ii) along the depth of the RS filter (C1-C5, as displayed in Figure 3.2 plot d), (iii) across parallel RS filters within a filter gallery and (iv) across filter galleries as well as temporal (monthly/seasonal) groupings for both treatment plants. A significant ANOVA finding was further investigated by preforming a post-hoc Tukey honest significant differences (HSD) test in R using the function Tukey HSD provided in the stats package.

Principal coordinate ordination plots was constructed using the first two axes from structure-based dissimilarity measures, Bray-Curtis and weighted UniFrac; and membership-based dissimilarity measures, Jaccard and unweighted UniFrac to visually represent pair-wise distances between samples in a multidimensional space. Pairwise analysis of molecular variance (AMOVA) (Excoffier *et al*., 1992) and parsimony comparisons using structure-based and membership-based dissimilarity measures were performed to test for significant differences between spatial and monthly groupings.

Distance-decay relationships were determined to assess the correlation between the change in microbial community membership and structure over distance along the surface and depth of the RS filter bed. Linear regression analysis was performed in R using functions lm and cor.test provided in the stats package to evaluate the correlation between the microbial community membership and structure, and distance. Scatter plots were constructed in R using the function ggplot with incorporated parameters geom_point and geom_jitter provided in the ggplot2 package (Wickham, 2016).

3.3 RESULTS

3.3.1 Drinking Water Treatment Plant 1

Sequence data

A total of 5 392 951 (M ± SD, 28 686 ± 8 063) raw V4 16S rRNA gene sequences were generated from the DNA extracts of 188 samples (Table A2). The sequences of two samples were removed from the raw sequence database before processing, as these sequences resulted from failed sequencing runs. Processing of the sequences following quality and chimera filtering removed approximately 19% of the sequences in the database. The final sequence database consisted of 4 374 697 processed sequences with an average read length of 250 base pairs. The average number of processed sequences per sample was 23 270 ± 6 327, with a minimum and maximum number of 2 050 and 41 614 sequences per sample, respectively (Table 3.1). To determine whether enough sequences were generated per sample to accurately represent the total community, rarefaction curves were constructed based on the number of observed microbial operational taxonomic units (OTUs). Most of the rarefaction curves (Figure A1) did not reach a plateau, indicating that the number of microbial OTUs observed was not entirely representative of the true microbial community.

Table 3.1 Summary of the raw and processed hypervariable V4 region of the 16S rRNA gene sequences that were generated using the Illumina MiSeq sequencing platform

Sequence data criteria	
Total number raw sequences	5 392 951
Average raw sequences	28 686
Standard deviation raw sequences	8 063
Total number processed sequences	4 374 697
Average processed sequences	23 270
Standard deviation processed sequences	6 327
Maximum processed sequences	41 614
Minimum processed sequences	2 050

Microbial diversity

To retain all of the samples for further comparisons, and at the same time ensure that the samples were compared at the same sequence depth during alpha diversity analysis, each sample was subsampled a 1 000 times to the sample containing the least amount of sequences (n_{seqs} = 2 050). To assess whether enough sequences per samples were retained after subsampling, Good's coverage scores were calculated based on 1 000 iterations (MOTHUR software, function: summary.single). The Good's coverage scores (Table A2) were high for all samples, ranging from 98.90% to 99.90%, indicating that the sequences retained within each sample after subsampling are adequate to describe the microbial communities reliably.

The microbial communities inhabiting the RS filters were diverse throughout the study period (Table A3). However, there were no significant differences in the alpha diversity indexes (species observed (S_{obs}), Shannon index (H') and Pielou's evenness index (J)) amongst the spatial groupings: (i) across the surface of the RS filter bed (one-way ANOVA, all $p > 0.05$, F_{ST} = 0.53, 0.49 and 0.30 for S_{obs}, H' and J, respectively) (Table A4), (ii) along the depth of the RS filter bed (one-way ANOVA, all $p > 0.05$, F_{ST} = 1.56, 2.38 and 1.15 for S_{obs}, H' and J, respectively) (Table A5), and (iii) across parallel RS filters within the same filter gallery [PS1 (one-way ANOVA, all $p > 0.05$, F_{ST} = 0.17, 0.30 and 0.69 for S_{obs}, H' and J, respectively), PS3 (one-way ANOVA, all $p > 0.05$, F_{ST} = 1.24, 0.24 and 1.05 for S_{obs}, H' and J, respectively) and PS4 (one-way ANOVA, all $p > 0.05$, F_{ST} = 0.85, 2.58 and 1.85 for S_{obs}, H' and J, respectively)] (Table A6).

Although no significant differences were observed when samples were grouped spatially along the depth of the RS filter bed, decreasing patterns in S_{obs} and H' were observed, while no decreasing pattern in J was noted. As shown in Figure 3.3 the microbial communities of samples associated with the two upper layers of the RS filter bed – C1 and C2 were more rich (S_{obs}, M ± SD = 249 ± 125 and 241 ± 196 for C1 and C2, respectively) and more diverse (H', M ± SD = 2.08 ± 0.35 and 1.94 ± 0.42 for C1 and C2, respectively), when compared to the samples associated with the lower depths of the RS filter bed [C3 (M ± SD = 198 ± 122 and 1.81 ± 0.53 for S_{obs} and H', respectively), C4 (M ± SD = 186 ± 154 and 1.60 ± 0.61 for S_{obs} and H', respectively) and C5 (M ± SD = 120 ± 77 and 1.58 ± 0.40 for S_{obs} and H', respectively)].

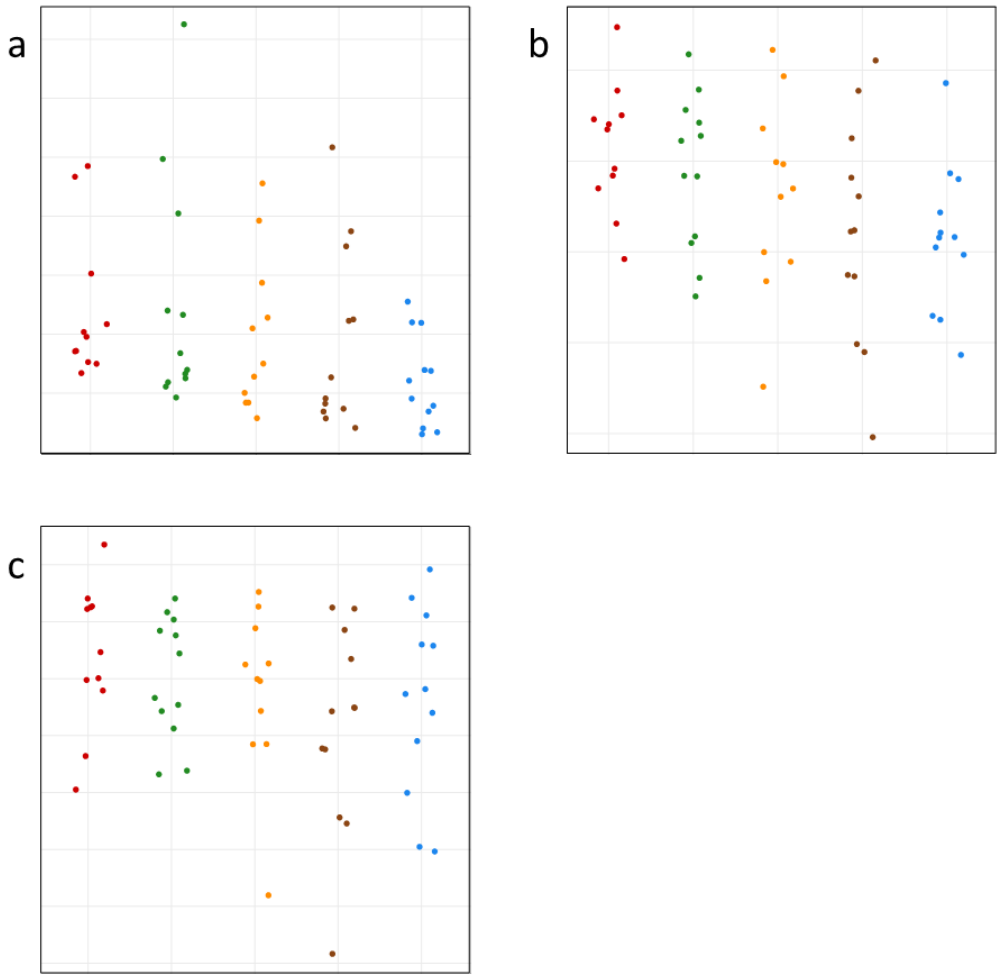


Figure 3.3 Alpha diversity indexes (a) observed species (Sobs), (b) Shannon index (H') and (c) Pielou's evenness index (J) for samples affiliated to spatial groupings: core location 1 (C1, red, nsamples = 11), core location 2 (C2, green, nsamples = 12), core location 3 (C3, orange, nsamples = 11), core location 4 (C4, brown, nsamples = 12) and core location 5 (C5, blue, nsamples = 12). See Table A5 for means, standard deviations and ANOVA tables.

In contrast, significant differences in the alpha diversity indexes were found for the spatial groupings across the different filter galleries (one-way ANOVA, all $p < 0.05$, F_{ST} = 59.71, 12.98 and 7.24 for S_{obs}, H' and J, respectively) (Table A7). As shown in Figure 3.4 the microbial communities of samples associated with PS1 were significantly more rich (S_{obs}, M ± SD = 322 ± 127) when compared to PS3 (S_{obs}, M ± SD = 156 ± 78) and PS4 (S_{obs}, M ± SD = 175 ± 62) (Tukey's post-hoc HSD test, all $p < 0.05$). Similarly, the microbial communities of samples associated with PS4 were significantly richer when compared to PS3 (Tukey's post-hoc HSD test, $p < 0.05$). Moreover, the microbial communities of samples associated with PS1 and PS4 were significantly more diverse (H', M ± SD = 2.15 ± 0.33 and 2.09 ± 0.38 for PS1 and PS4, respectively) when compared to PS3 (H', M ± SD = 1.81 ± 0.45) (Tukey's post-hoc HSD test, all $p < 0.05$); while the microbial communities of samples associated with PS4 were significantly more even (J, M ± SD = 0.39 ± 0.06) when compared to PS1 (J, M ± SD = 0.36 ± 0.05) and PS3 (J, M ± SD = 0.35 ± 0.07) (Tukey's post-hoc HSD test,

all $p < 0.05$). Neither the diversity (H') between PS1 and PS4, nor the evenness (J) between PS1 and PS3 were significantly different (Tukey's post-hoc HSD test, all $p > 0.05$).

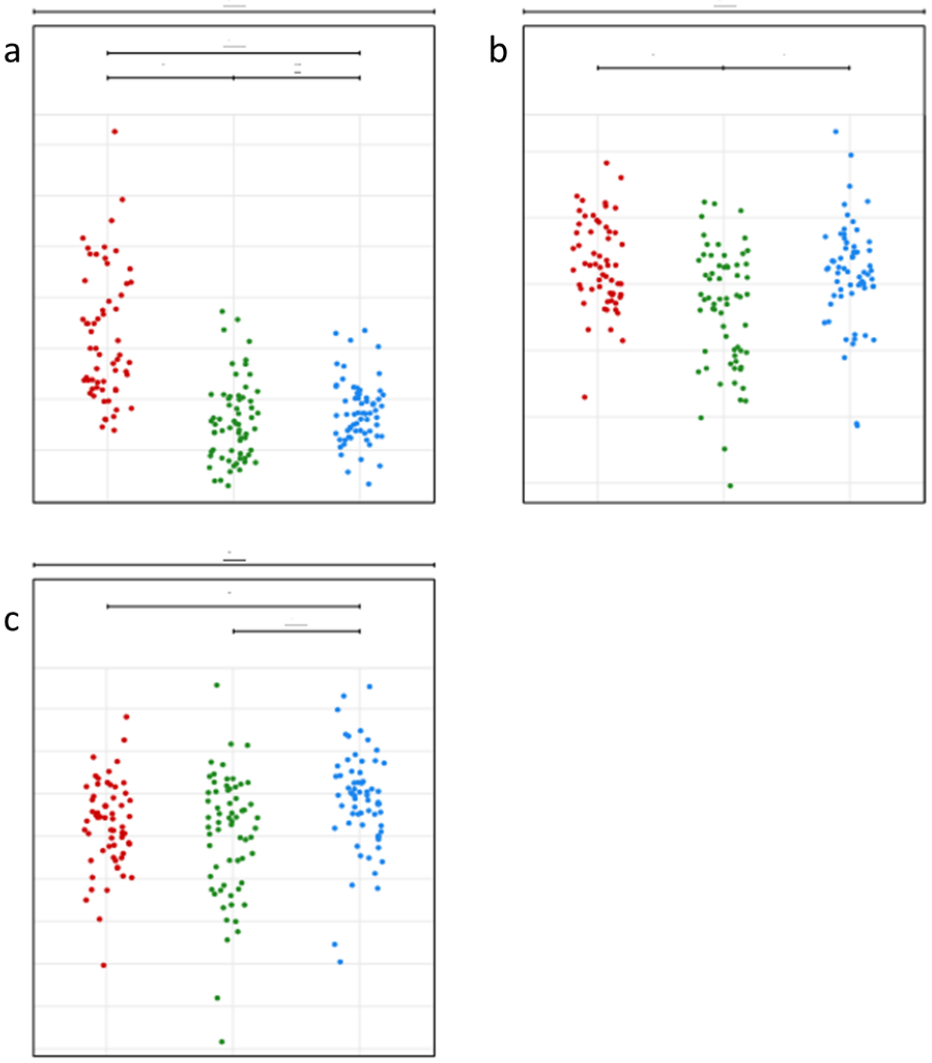


Figure 3.4 Alpha diversity indexes (a) observed species (S_{obs}), (b) Shannon index (H') and (c) Pielou's evenness index (J) for samples affiliated to spatial groupings: purification station 1 (PS1, red, $n_{samples}$ = 60), purification station 3 (C2, green, $n_{samples}$ = 64) and purification station 4 (C3, blue, $n_{samples}$ = 64). All significant values are indicated with bars at the top of the figures (p values: * = $0.01 < p \leq 0.05$, ** = $0.001 < p \leq 0.01$, *** = $0.0001 < p \leq 0.001$, **** = $p \leq 0.001$). See Table A7 for means, standard deviations, ANOVA tables and post-hoc Tukey HSD tests.

Furthermore, significant differences in temporal groupings were found when samples were grouped based on month [PS1 (one-way ANOVA, all $p < 0.05$, F_{ST} = 8.31, 12.73 and 19.02 for S_{obs}, H' and J, respectively) (Table A8), PS3 (one-way ANOVA, all $p < 0.05$, F_{ST} = 3.02, 5.45 and 4.26 for S_{obs}, H' and J, respectively) (Table A9) and PS4 (one-way ANOVA, all $p < 0.05$, F_{ST} = 3.17 and 3.40 for H' and J, respectively) (Table A10)]. At PS1, the microbial communities of samples collected in May and August were significantly richer (S_{obs}, M ± SD = 409 ± 132 and 363 ± 129 for May and August, respectively) when compared to June (S_{obs}, M ± SD = 257 ± 72) and July (S_{obs}, M ± SD = 243 ± 85) (Tukey's post-hoc HSD test, all $p < 0.05$). Moreover, at

PS1 the microbial communities of the samples collected in May were significantly more diverse (H', M ± SD = 2.44 ± 0.29) and even (J, M ± SD = 0.40 ± 0.03) when compared to the microbial communities of the samples collected in June (M ± SD = 2.10 ± 0.17 and 0.35 ± 0.02 for H' and J, respectively) (Tukey's post-hoc HSD test, all $p < 0.05$). In contrast, at PS1 the microbial communities of samples collected during August were significantly less diverse (H', M ± SD = 1.91 ± 0.31) when compared to May (H', M ± SD = 2.44 ± 0.29) and July (H', M ± SD = 2.22 ± 0.25) (Tukey's post-hoc HSD test, all $p < 0.05$), and less even (J, M ± SD = 0.31 ± 0.05) when compared to May (J', M ± SD = 0.40 ± 0.03), June (J', M ± SD = 0.35 ± 0.02) and July (J', M ± SD = 0.40 ± 0.04) (Tukey's post-hoc HSD test, all $p < 0.05$).

In contrast to PS1, at PS3 the microbial communities of samples collected during August were significantly richer (S_{obs}**,** M ± SD **=** 197 ± 110), more diverse (H', M ± SD **=** 2.07 ± 0.38), and more even (J, M ± SD = 0.39 ± 0.04) when compared to May (M ± SD = 118 ± 41, 1.50 ± 0.33 and 0.31 ± 0.05 for S_{obs}, H' and J, respectively) (Tukey's post-hoc HSD test, all $p < 0.05$). Additionally, at PS3 the microbial communities of samples collected during May were significantly less diverse (H', M ± SD = 1.50 ± 0.33) and less even (J, M ± SD = 0.31 ± 0.05) when compared to June (M ± SD = 1.90 ± 0.35 and 0.37 ± 0.07 for H' and J, respectively) (Tukey's post-hoc HSD test, all $p < 0.05$). Similar to PS1, at PS4 the microbial communities of samples collected during May were significantly more diverse (H', M ± SD = 2.29 ± 0.46) when compared to July (H', M ± SD = 1.93 ± 0.48) (Tukey's post-hoc HSD test, $p < 0.05$), and more even (J, M ± SD = 0.42 ± 0.07) when compared to June (J, M ± SD = 0.37 ± 0.04) and July (J M ± SD = 0.37 ± 0.07) (Tukey's post-hoc HSD test, $p < 0.05$).

Microbial community composition

Classification of the 4 374 697 processed sequences into OTUs, defined at a 97% sequence similarity threshold, identified 5 708 microbial OTUs. Of these OTUs, 5 564 were bacterial and constituted 99.91% of the total sequences (n_{seqs} = 4 370 872), while the remaining 144 were classified as archaeal OTUs constituting 0.09% of the total sequences (n_{seqs} = 3 825). To describe the microbial communities, the OTUs were split into their taxonomic ranks (i.e. phyla, class, order, family and genus) and grouped as either dominant or rare using a relative abundance (RA) threshold of 1% as described by Gobet *et al*. (2010). At phyla level, 61 phyla were identified, comprising of 3 archaeal phyla and 58 bacterial phyla. Amongst these microbial phyla, 3 dominant phyla were identified that collectively accounted for approximately 98.19% of the total sequences (n_{seqs} = 4 294 362). These phyla, averaged across all study sites, in decreasing order were: *Proteobacteria* (mean relative abundance, MRA = 86.44%, n_{seqs} = 3 781 596), *Bacteroidetes* (MRA = 9.88%, n_{seqs} = 432 336) and *Actinobacteria* (MRA = 1.84%, n_{seqs} = 80 430) (Figure 3.5a).

Within *Proteobacteria*, the sequences were mainly distributed across *Beta*- and *Gammaproteobacteria* (Figure 3.5b). The proteobacterial sequences that were classified as *Gammaproteobacteria,* on average, accounted for 90.25% of the proteobacterial sequences (n_{seqs} = 3 412 985); flowed by *Betaproteobacteria* (MRA = 8.54%, n_{seqs} = 322 885). Of the *Gammaproteobacteria* sequences, approximately 44.93% (n_{seqs} = 1 533 563), 42.43% (n_{seqs} = 1 447 967) and 10.84% (n_{seqs} = 370 199) were affiliated to *Pseudomonadales*, *Aeromonadales* and *Alteromonadales*, respectively. Within these dominant families some OTUs were further

classified to genus level, i.e. the genera *Acinetobacter* and *Pseudomonas* within *Pseudomonadales,* the genus *Aeromonas* within *Aeromonadales* and *Shewanella* and *Rheinheimera* within the family *Alteromonadales*. Amongst these genera, *Aeromonas* and *Acinetobacter* were the two most commonly detected genera in the FB samples, and accounted for 33.11% (n_{seqs} = 1 447 949) and 31.08% (n_{seqs} = 1 359 477) of the total sequences, respectively; while *Shewanella*, *Pseudomonas* and *Rheinheimera* accounted for 6.47% (n_{seqs} = 282 925), 3.87% (n_{seqs} = 169 339) and 1.99% (n_{seqs} = 87 045) of the total sequences, respectively. Of the *Betaproteobacteria* sequences, the majority, that is 60.10% (n_{seqs} = 203 645), was affiliated to *Burkholderiales*, with 54.15% (n_{seqs} = 110 272) and 45.85% (n_{seqs} = 93 373) of the *Burkholderiales* sequences belonging to *Janthinobacterium* and *Comamonas*, respectively. In addition to *Proteobacteria*, sequences of *Bacteroidetes*, which was the second most dominant phylum in the FB samples, was primarily classified as *Flavobacterium* and accounted for 9.35% of the total sequences (n_{seqs} = 408 970). The third most dominant phylum, *Actinobacteria* was mainly affiliated to the ACK-M1 clade that accounted for 1.30% of the total sequences (n_{seqs} = 56 746).

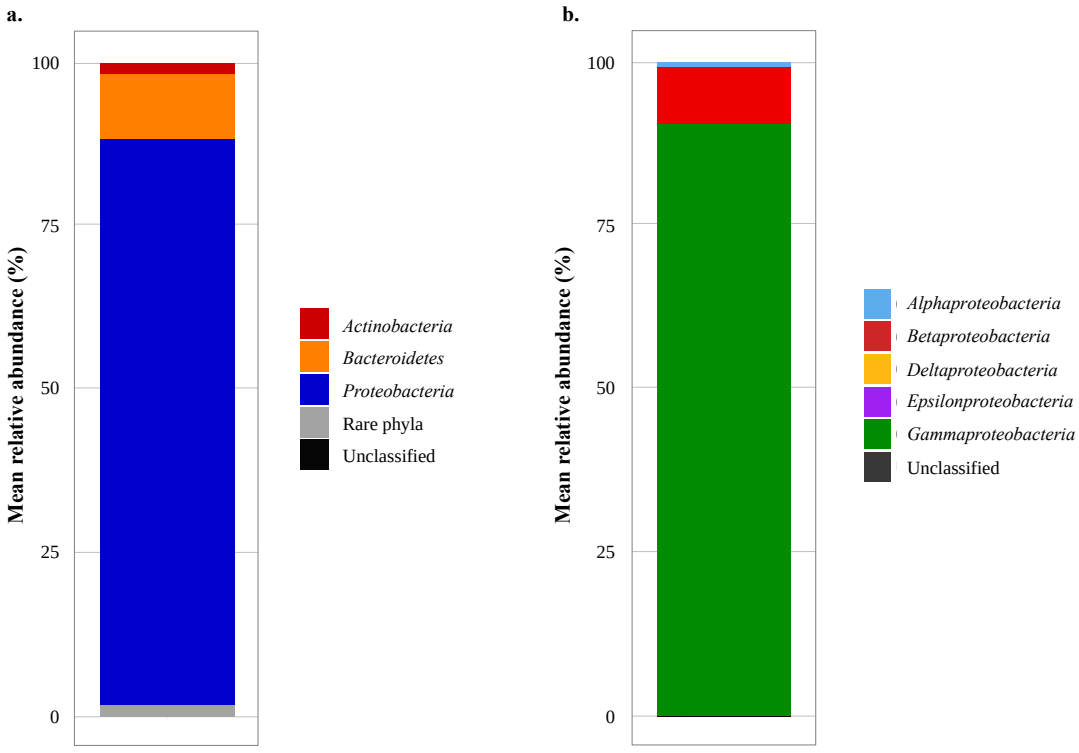


Figure 3.5 (a) Stacked bar charts displaying the mean relative abundance of the (a) three dominant microbial phyla recovered from the rapid sand filters. (b) Stacked bar charts displaying the mean relative abundance of the *Proteobacteria* subphyla recovered from the study sites.

Spatial and temporal variations of the rapid sand filter microbial communities

Pairwise analysis of molecular variance (AMOVA) and parsimony comparisons using membership- and structure-based dissimilarity measures (i.e. Jaccard (d_J), unweighted UniFrac (d_{UUF}) and Bray-Curtis (d_{BC}), weighted UniFrac (d_{WUF}), respectively) were preformed to test for significant differences between spatial and

monthly groupings. Following these analyses, it was shown that there were no significant differences in the microbial community membership or structure when samples were grouped spatially: (i) across the surface of the RS filter bed (AMOVA, all $p > 0.05$, $F_{ST} = 0.33, 0.34, 0.80, 0.78$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p > 0.05$) (Table A11), (ii) along the depth of the RS filter bed (AMOVA, all $p > 0.05$, $F_{ST} = 0.69, 0.73, 1.13, 1.15$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p > 0.05$) (Table A12), and (iii) across parallel RS filters within the same filter gallery [PS1 (AMOVA, all $p > 0.05$, $F_{ST} = 0.61, 0.91, 0.98, 0.95$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p > 0.05$), PS3 (AMOVA, all $p > 0.05$, $F_{ST} = 0.78, 0.65, 0.99, 1.00$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p > 0.05$) and PS4 (AMOVA, all $p > 0.05$, $F_{ST} = 0.58, 0.80, 1.02, 0.98$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p > 0.05$) (Table A13)].

In contrast, significant differences in the microbial community membership and structure were found amongst the spatial groupings across the different filter galleries (AMOVA, all $p < 0.05$, $F_{ST} = 2.44, 2.04, 2.05, 2.00$ for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively; parsimony, all $p < 0.05$) (Table A14). In particular, post-hoc pairwise AMOVA tests revealed that the microbial community memberships and structures between spatial groupings PS1-PS3, and PS1-PS4 were significantly different (all $p < 0.05$); however the microbial community memberships and structures between spatial grouping PS3-PS4 were not significantly different (all $p > 0.05$).

To visually display the spatial variability in the microbial community membership and structure amongst spatial groupings across the different filter galleries, principal coordinate (PCoA) ordination plots were constructed on OTU level using the first two axes of the membership- and structure-based dissimilarity measures (Figure 3.6). The PCoA plots showed clustering of the samples according to individual spatial groupings within each purification system. These clusters were more apparent in the membership-based PCoA plots (Figures 3.6c and 3.6d) compared to the structure-based PCoA plots (Figures 3.6a and 3.6b). From the structure-based PCoA plots, no discrepancy amongst samples of spatial groupings PS1, PS3 and PS4 were observed, as these samples were randomly distributed along the first and second PCoA axes (Figures 3.6a and 3.6b). Although clustering was found, the microbial community structures of these spatial groupings were between 30% and 50% dissimilar, depending on the structure-based dissimilarity measure used and the spatial groupings being compared (Table 3.2). As shown in Figures 3.7a and 3.7b, the averaged dissimilarities in the microbial community structures were slightly higher between spatial groupings PS1 vs. PS3 ($M \pm SD = 0.52 \pm 0.14$ and 0.36 ± 0.08 for d_{BC} and d_{WUF}, respectively) and PS1 vs. PS4 ($M \pm SD = 0.49 \pm 0.11$ and 0.35 ± 0.09 for d_{BC} and d_{WUF}, respectively), when compared to PS3 vs. PS4 ($M \pm SD = 0.47 \pm 0.18$ and 0.33 ± 0.12 for d_{BC} and d_{WUF}, respectively).

From the membership-based PCoA plots, the samples from PS1 clearly clustered together and away from samples of PS3 and PS4 (Figures 3.6c and 3.6d). Specifically, samples of spatial grouping PS1 clustered independently along the first PCoA axis, whereas samples of spatial groupings PS3 and PS4 clustered along the second PCoA axis. As shown in Figures 3.7c and 3.7d significant higher averaged dissimilarities in the microbial community memberships were found between spatial groupings PS1 vs. PS3 ($M \pm SD = 0.80 \pm 0.04$ and 0.73 ± 0.05 for d_J and d_{UUF}, respectively), followed by PS1 vs. PS4 ($M \pm SD = 0.78 \pm 0.04$ and 0.72

± 0.05 for d_J and d_{UUF}, respectively), and PS3 vs. PS4 (M ± SD = 0.73 ± 0.04 and 0.68 ± 0.05 for d_J and d_{UUF}, respectively) (ANOVA, all $p < 0.05$, F_{ST} = 99.94 and 43.44 for d_J and d_{UUF}, respectively) (Table A15).

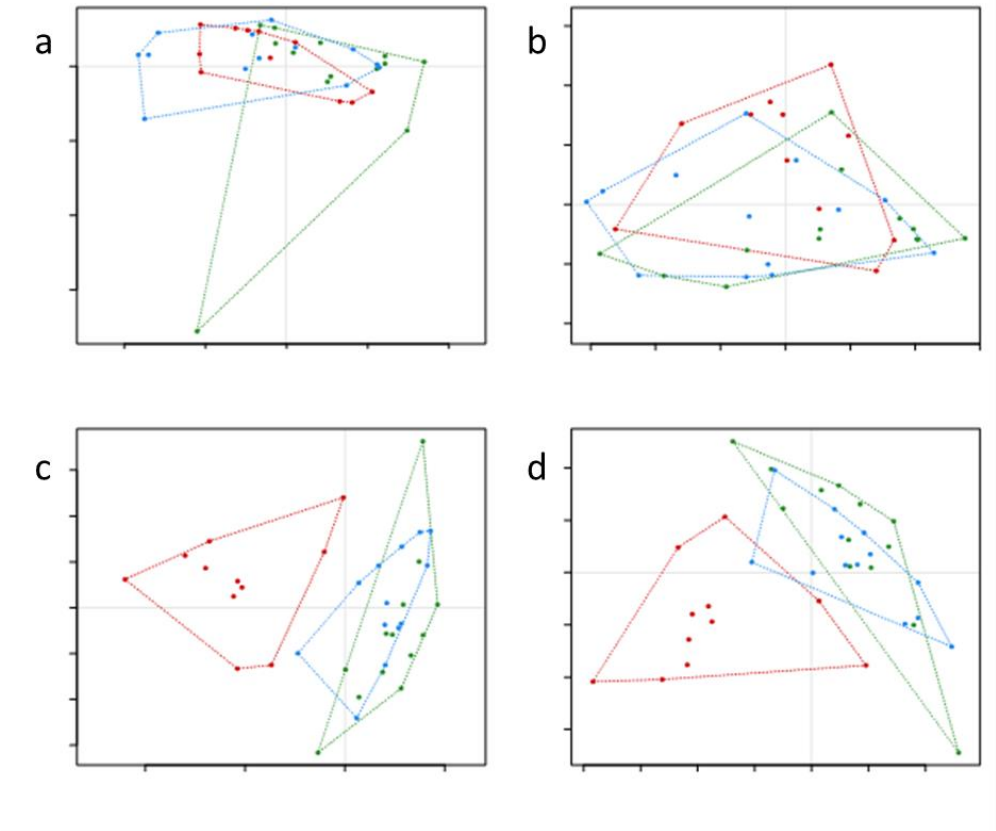


Figure 3.6 Principal coordinate ordination of the RS filter data showing the distribution of purification station 1 (red), purification station 3 (green) and purification station 4 (blue) samples in the first two axis of a multidimensional space using (a) Bray-Curtis, (b) weighted UniFrac, (c) Jaccard and (d) unweighted UniFrac distances. The percentages on the axis represents the variance explained by each of the coordinates.

Table 3.2 Means and standard deviations of structure and membership-base distances between purification stations in the drinking water treatment plant

	Structure-based diversity matrices				**Membership-based diversity matrices**			
	Bray-Curtis		**Weighted UniFrac**		**Jaccard**		**Unweighted UniFrac**	
Purification station	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
PS1 vs. PS3	0.52	0.14	0.36	0.08	0.8	0.04	0.73	0.05
PS1 vs. PS4	0.49	0.11	0.35	0.09	0.78	0.04	0.72	0.05
PS3 vs. PS4	0.47	0.18	0.33	0.12	0.73	0.04	0.68	0.05

Abbreviations: **PS1**, purification station 1; **PS3**, purification station 3; **PS4**, purification station 4; **n**, number of samples; **M**, mean; **SD**, standard deviation.

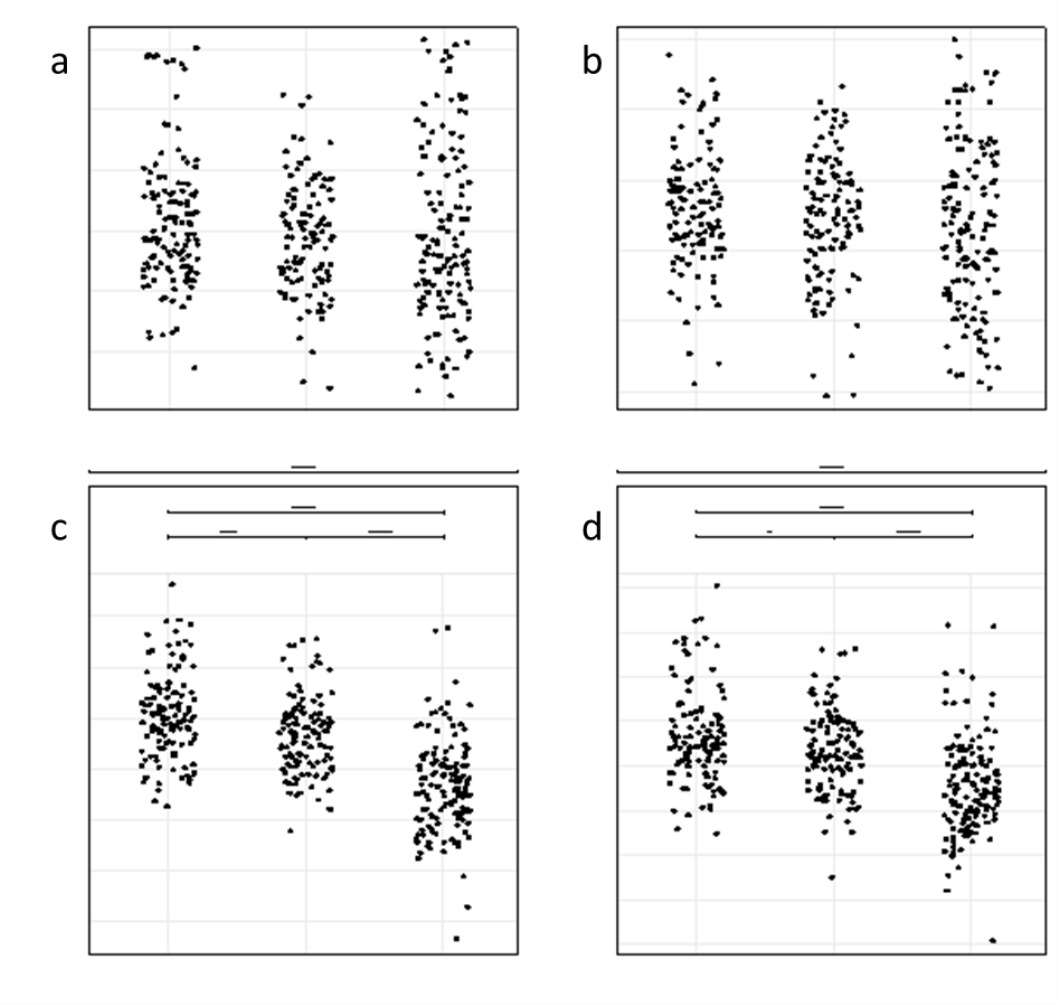


Figure 3.7 Pairwise (a) Bray-Curtis, (b) weighted UniFrac, (c) Jaccard and (d) unweighted UniFrac distances between samples of spatial groupings: (i) PS1 vs. PS3 ($n_{samples}$ = 120), (ii) PS1 vs. PS4 ($n_{samples}$ = 120) and (iii) PS3 vs. PS4 ($n_{samples}$ = 144). All significant values are indicated with bars at the top of the figures (p values: * = 0.01 < p ≤ 0.05, ** = 0.001 < p ≤ 0.01, *** = 0.0001 < p ≤ 0.001, **** = p ≤ 0.001). See Table B3.15 for means, standard deviations, ANOVA tables and post hoc Tukey HSD tests.

Moreover, the high dissimilarity in the microbial community membership mentioned above (i.e. between 70% and 80% dissimilarity between purification systems), depending on the membership-based dissimilarity measured used, was in large accredited to numerous low-abundant OTUs that were exclusively associated with each purification system. In particular, as shown in Figure 3.8 approximately 38.77% (n_{OTUs} = 2 213), 17.16% (n_{OTUs} = 979) and 16.12% (n_{OTUs} = 920) of the total OTUs were exclusively associated with samples of the PS1, PS3 and PS4, respectively, where they only accounted for between 0.04% and 0.37% of the total sequences.

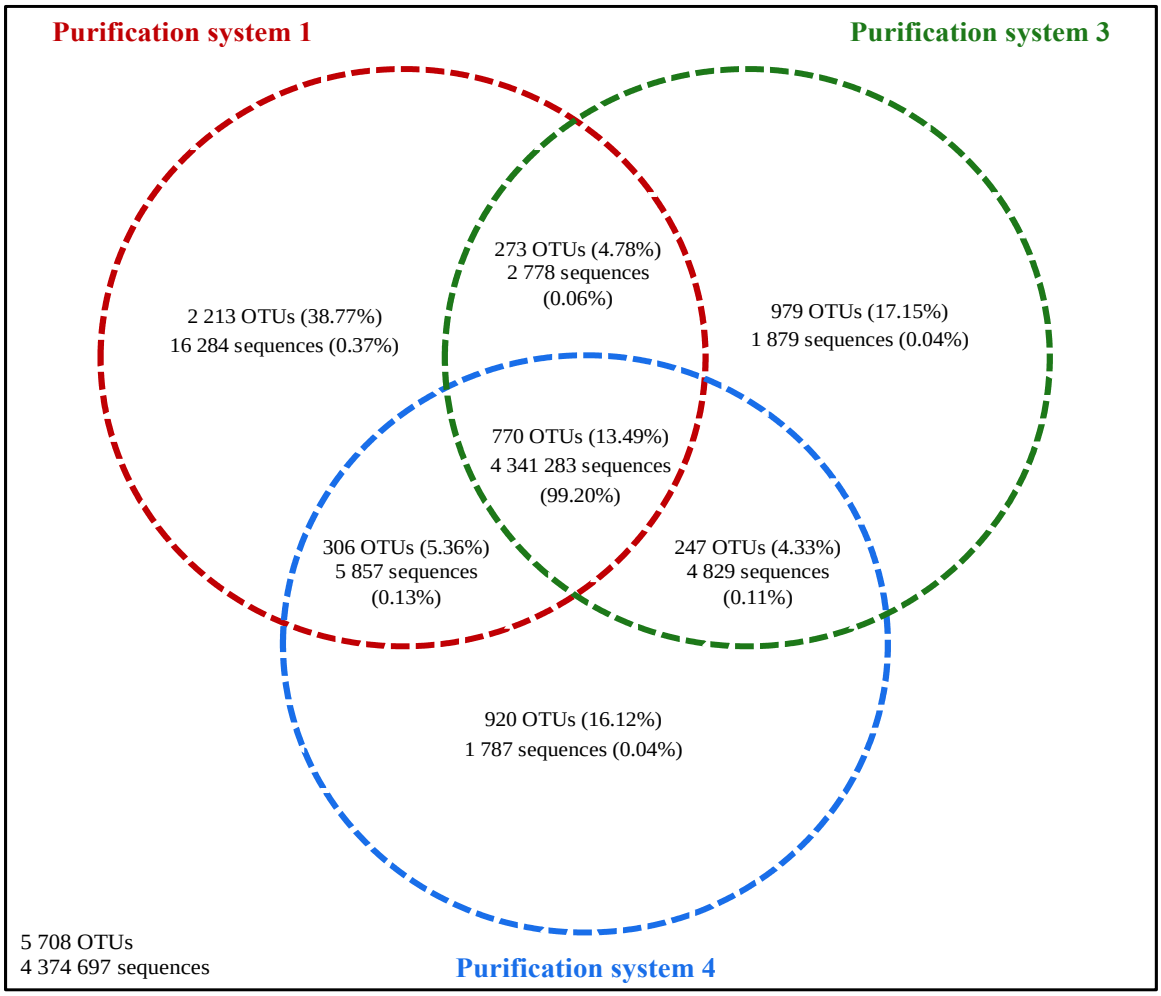


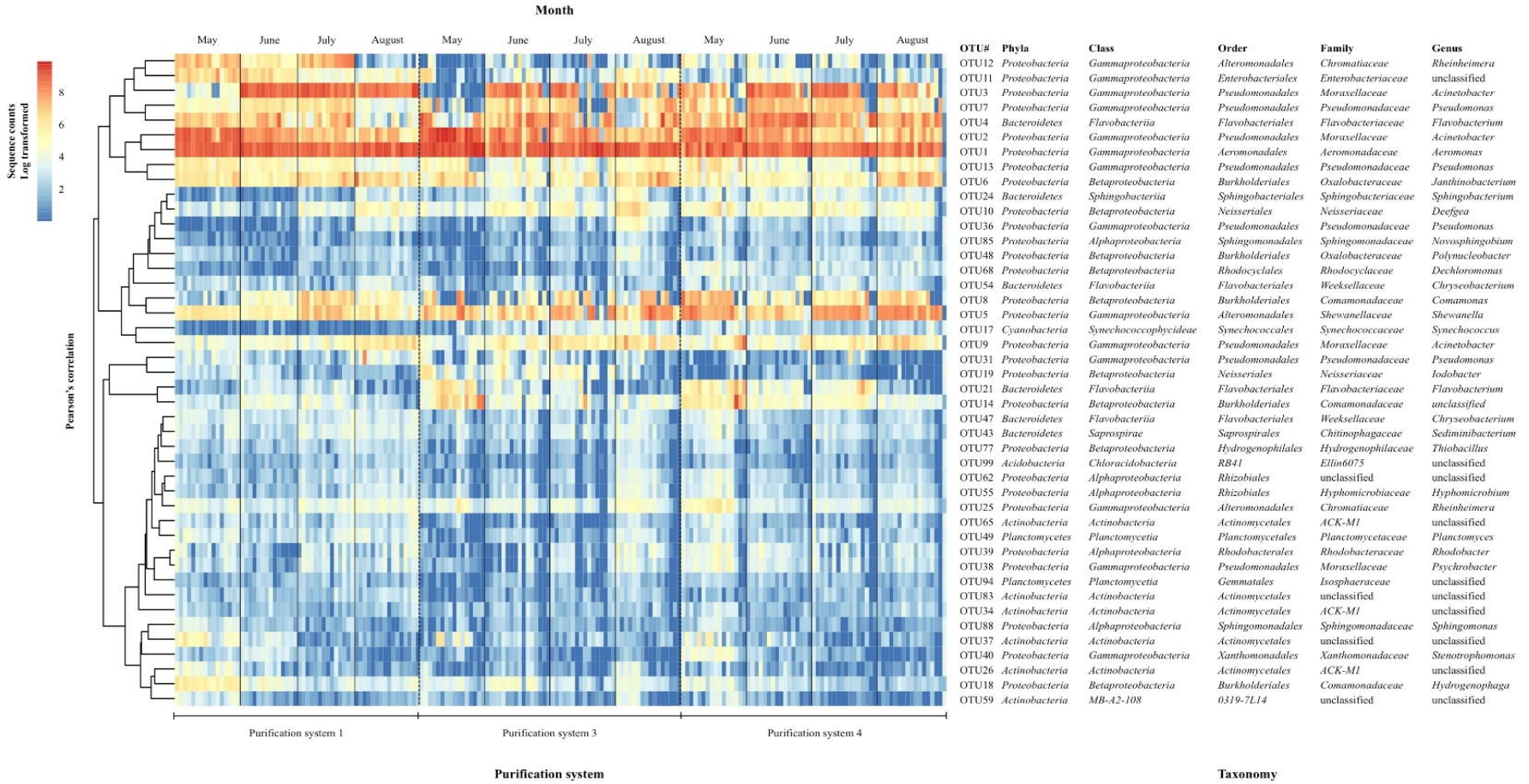
Figure 3.8 Venn diagrams displaying the shared and unshared OTUs in terms of counts and sequence abundances detected amongst purification station 1 (red), purification station 3 (green) and purification station 4 (blue).

Temporal variations of the rapid sand filter microbial communities

In addition to the spatial variability, temporal variation within each of the purification system was displayed in the PCoA plots (Figure 3.6). From the membership- and structure-based PCoA plots, no apparent temporal dispersion patterns were observed amongst spatial groupings PS1, PS3 and PS4. The test of homogeneity of multivariate dispersion provided strong evidence to support that no differences in temporal dispersion amongst the spatial groupings occurred (ANOVA, all $p > 0.05$, F_{ST} = 0.41, 2.51, 0.36 and 0.85 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively) (Table A16). Specifically, the average distance to centroid was not significantly larger for samples of spatial grouping PS1 ($\bar{z}$ = 0.29, 0.21, 0.51 and 0.46 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively), PS3 ($\bar{z}$ = 0.33, 0.23, 0.50 and 0.46 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively), and PS4 ($\bar{z}$ = 0.32, 0.22, 0.48 and 0.44 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively). Although similar temporal variations in the microbial communities of the PS spatial groupings were observed, significant differences in the microbial community memberships and structures were found amongst the month groupings for PS1 (AMOVA, all $p <$ 0.05, F_{ST} = 4.38, 4.12, 1.35, 1.46 for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively) (Table A17), PS3 (AMOVA, all $p <$

0.05, F_{ST} = 1.58, 0.70, 1.16, 1.13 for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively) (Table A18), and PS4 (AMOVA, all $p < 0.05$, F_{ST} = 2.63, 3.53, 1.29, 1.21 for d_{BC}, d_{WUF}, d_J and d_{UUF} respectively) (Table A19).

The spatial and temporal trends of selected microbial OTUs in the RS filters of the filter galleries were visually represented in a heatmap that was constructed using the log transformed sequence abundances of 44 OTUs (Figure 3.9). These OTUs were selected based on their relative abundance within a sample (i.e. relative abundance threshold ≥ 1%) and by the percentage of FB samples in which they were detected (i.e. detection frequency threshold ≥ 75%). Although this subset of OTUs accounted for a notable small proportion ~ 0.77% of the original OTU dataset, it accounted for approximately 95% of the overall relative abundances and, in addition, was shown to sufficiently explain the variation observed in the microbial community structure (Mantel's test, all $p \leq 0.05$; $r_{\text{Bray-Curtis}}$ = 0.993, $r_{\text{weighted UniFrac}}$ = 0.994). At phyla level, these OTUs were distributed across 6 phyla. These phyla, averaged across the 188 FB samples, in decreasing order were: *Proteobacteria* (MRA = 89.25%, n_{seqs} = 3 706 102, n_{OTUs} = 28), *Bacteroidetes* (MRA = 10.14%, n_{seqs} = 420 897, n_{OTUs} = 6), *Actinobacteria* (MRA = 0.40%, n_{seqs} = 16 811, n_{OTUs} = 6), *Planctomycetes* (MRA = 0.11%, $n_{sequences}$ = 4 665, n_{OTUs} = 2), *Cyanobacteria* (MRA = 0.07%, n_{seqs} =2 898, n_{OTU} = 1) and *Acidobacteria* (MR*A* = 0.03%, n_{seqs} = 1 231, n_{OTU} = 1).



OTU#	Phyla	Class	Order	Family	Genus
OTU12	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Chromatiaceae*	*Rheinheimera*
OTU11	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	unclassified
OTU3	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter*
OTU7	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas*
OTU4	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacterium*
OTU2	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter*
OTU1	*Proteobacteria*	*Gammaproteobacteria*	*Aeromonadales*	*Aeromonadaceae*	*Aeromonas*
OTU13	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas*
OTU6	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	*Janthinobacterium*
OTU24	*Bacteroidetes*	*Sphingobacteriia*	*Sphingobacteriales*	*Sphingobacteriaceae*	*Sphingobacterium*
OTU10	*Proteobacteria*	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*	*Deefgea*
OTU36	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas*
OTU85	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Novosphingobium*
OTU48	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	*Polynucleobacter*
OTU68	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	*Dechloromonas*
OTU54	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Weeksellaceae*	*Chryseobacterium*
OTU8	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Comamonas*
OTU5	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella*
OTU17	*Cyanobacteria*	*Synechococcophycideae*	*Synechococcales*	*Synechococcaceae*	*Synechococcus*
OTU9	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter*
OTU31	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas*
OTU19	*Proteobacteria*	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*	*Iodobacter*
OTU21	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacterium*
OTU14	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	*unclassified*
OTU47	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Weeksellaceae*	*Chryseobacterium*
OTU43	*Bacteroidetes*	*Saprospirae*	*Saprospirales*	*Chitinophagaceae*	*Sediminibacterium*
OTU77	*Proteobacteria*	*Betaproteobacteria*	*Hydrogenophilales*	*Hydrogenophilaceae*	*Thiobacillus*
OTU99	*Acidobacteria*	*Chloracidobacteria*	*RB41*	*Ellin6075*	unclassified
OTU62	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	unclassified	unclassified
OTU55	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	*Hyphomicrobium*
OTU25	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Chromatiaceae*	*Rheinheimera*
OTU65	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*ACK-M1*	unclassified
OTU49	*Planctomycetes*	*Planctomycetia*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces*
OTU39	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Rhodobacter*
OTU38	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Psychrobacter*
OTU94	*Planctomycetes*	*Planctomycetia*	*Gemmatales*	*Isosphaeraceae*	unclassified
OTU83	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	unclassified	unclassified
OTU34	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*ACK-M1*	unclassified
OTU88	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingomonas*
OTU37	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	unclassified	unclassified
OTU40	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	*Stenotrophomonas*
OTU26	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*ACK-M1*	unclassified
OTU18	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Hydrogenophaga*
OTU59	*Actinobacteria*	*MB-A2-108*	*0319-7L14*	unclassified	unclassified

Taxonomy

Figure 3.9 Heatmap constructed using the absolute sequence abundances (log transformed) of 55 OTUs that were selected based on their relative abundance within a sample (i.e. relative abundance threshold ≥ 1%) and by the percentage of samples in which they were detected (i.e. detection frequency threshold ≥ 75%). The heatmap boxes were colored from red-to-blue to represent higher-to-lower abundances, and OTUs not represented by a sequence were assigned 10^{-6} (displayed as dark blue).

Four OTUs classified as unknown species of the genera *Aeromonas* (*Gammaproteobacteria*, OTU1), *Acinetobacter* (*Gammaproteobacteria*, OTU2 and OTU3) and *Flavobacterium* (*Bacteroidetes*, OTU4), which collectively accounted for about 75.17% (n_{seqs} = 3 121 634) of the total sequences, persistently dominated the microbial community of samples in spatial grouping PS1 [(OTU1: MRA = 36.00% (n_{seqs} = 466 430), OTU2: MRA = 15.01% (n_{seqs} = 194 524), OUT3: MRA = 23.50% (n_{seqs} = 304 466) and OTU4: MRA 4.51% (n_{seqs} = 58 493)], PS3 [(OTU1: MRA = 39.66% (n_{seqs} = 593 512), OTU2: MRA = 20.13% (n_{seqs} = 301 179), OTU3: MRA = 9.16% (n_{seqs} = 137 110) and OTU4: MRA = 8.61% (n_{seqs} = 128 813)] and PS4 [(OTU1: MRA = 28.50% (n_{seqs} = 387 751), OTU2: MRA = 12.04% (n_{seqs} = 163 767), OTU3: MRA = 14.12% (n_{seqs} = 192 176) and OTU4: MRA = 14.21% (n_{seqs} = 193 413)]. As inferred from the constructed phylogenetic trees, the unknown species of *Aeromonas* clustered with cultured relative *Aeromonas sorbia* (Figure 3.10), whereas the unknown species of *Flavobacterium* clustered close to *Flavobacterium succinicans* (Figures 3.12). For the two unknown *Acinetobacter* species, clustering were found with *Acinetobacter johnsonii*, and three known species *Acinetobacter bohemicus*, *Acinetobacter pakistanensis* and *Acinetobacter movanagherensis*. No base pair differences were observed for this short piece of the 16S rRNA sequence (Figure 3.11).

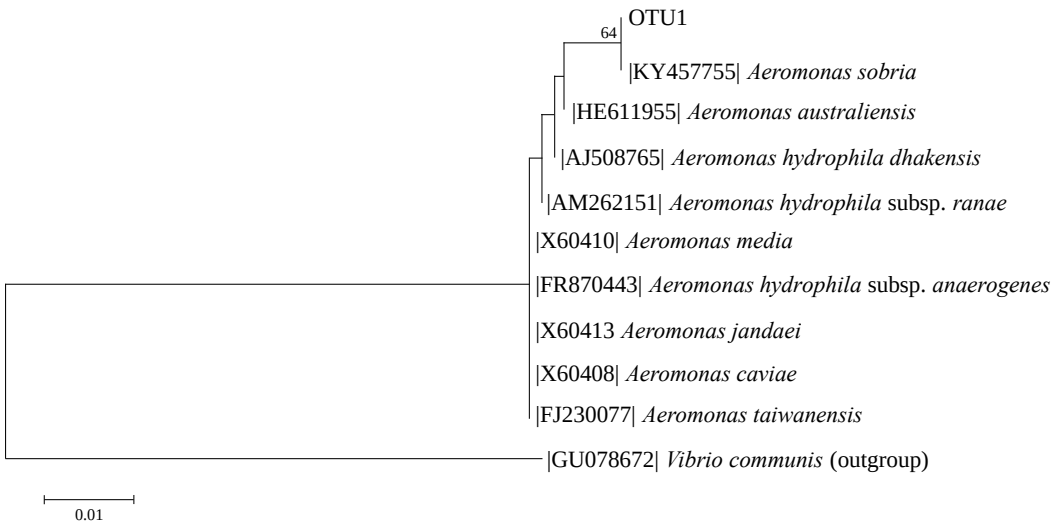


Figure 3.10 Neighbour joining phylogenetic tree showing the relationship between OTU1 and *Aeromonas* type sequences obtained from the NCBI GenBank® sequence database. The tree was rooted with *Vibrio communis* as the outgroup and bootstrap analysis of 1 000 replicates. Bootstrap values are indicated as percentages and values below 50 were excluded.

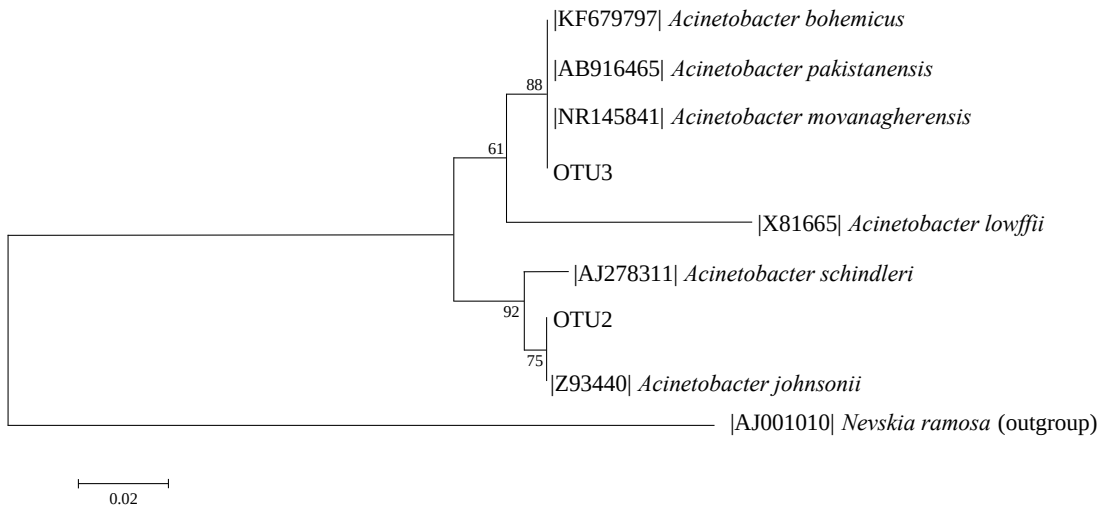


Figure 3.11 Neighbour joining phylogenetic tree showing the relationship between OTU2 and OTU3, and *Acinetobacter* type sequences obtained from the NCBI GenBank® sequence database. The tree was rooted with *Nevskia ramosa* as the outgroup and bootstrap analysis of 1 000 replicates. Bootstrap values are indicated as percentages and values below 50 were excluded.

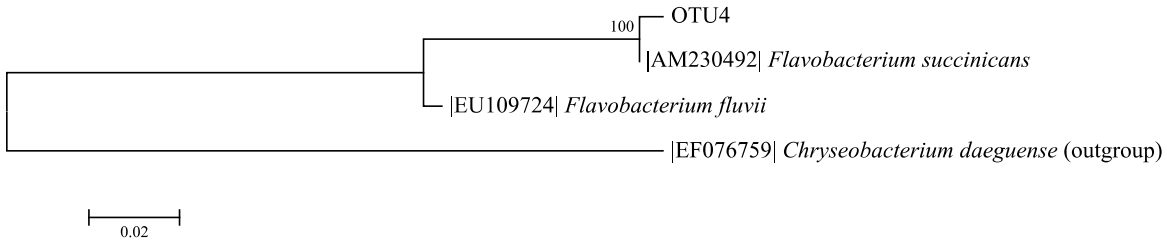


Figure 3.12 Neighbour joining phylogenetic tree showing the relationship between OTU4 and *Flavobacterium* type sequences obtained from the NCBI GenBank® sequence database. The tree was rooted with *Chryseobacterium daeguense* as the outgroup and bootstrap analysis of 1 000 replicates. Bootstrap values are indicated as percentages and values below 50 were excluded.

Moreover, two OTUs classified as unknown species of the family *Enterobacteriaceae* (*Gammaproteobacteria*, OTU11) and genera *Rheinheimera* (*Gammaproteobacteria*, OTU12) were detected at averaged relative abundances that were between 4-fold and 20-fold higher in spatial grouping PS1 (MRA = 2.16% (n_{seqs} = 28 036) and 4.61% (n_{seqs} = 59 688) for OTU11 and OTU12, respectively), when compared to spatial groupings PS3 (MRA = 0.55% (n_{seqs} = 8 181) and 0.22% (n_{seqs} = 3 320 for OTU11 and OTU12, respectively) and PS4 (MRA = 0.21% (n_{seqs} = 2 821) and 0.34% (n_{seqs} = 4 593) for OTU11 and OTU12, respectively). The unknown *Rheinheimera* species clustered with three known species *Rheinheimera tilapiae*, *Rheinheimera texasensis* and *Rheinheimera aquatica*. No base pair differences were observed for this short piece of the 16S rRNA sequence (Figures 3.13).

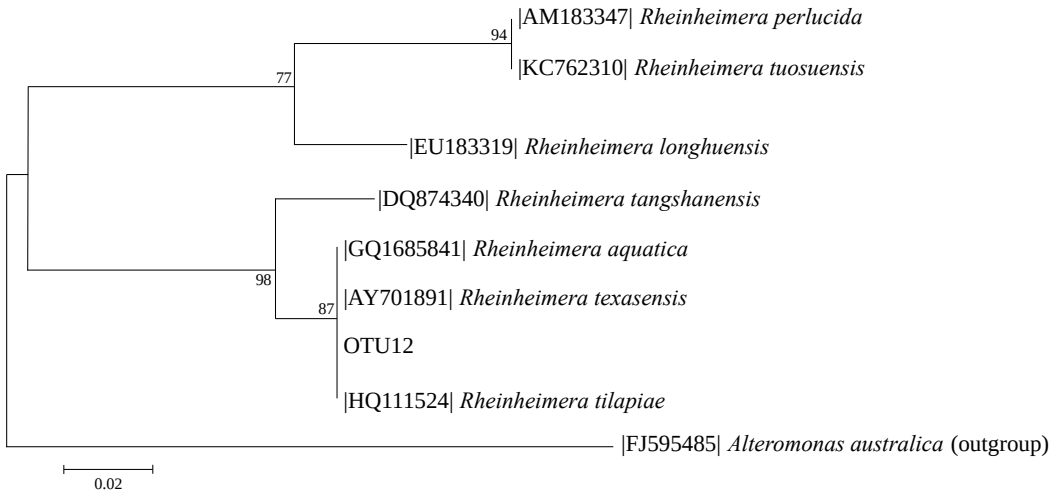


Figure 3.13 Neighbor joining phylogenetic tree showing the relationship between OTU12 and *Rheinheimera* type sequences obtained from the NCBI GenBank® sequence database. The tree was rooted with *Alteromonas australica* as the outgroup and bootstrap analysis of 1 000 replicates. Bootstrap values are indicated as percentages and values below 50 were excluded.

Local distance-decay relationships

Local distance-decay correlations were assessed using the distance between two locations along the surface and depth of the RS filter bed (see calculated distances Figure 3.2 plots a to c), and pair-wise membership- and structure-based dissimilarity measures (i.e. Jaccard (d_J), unweighted UniFrac (d_{UUF}) and Bray-Curtis (d_{BC}), weighted UniFrac (d_{WUF}), respectively). As shown in Figure 3.14, significant positive Pearson's correlations between distances across the surface of the RS filter bed and difference in the microbial community membership and structure were found (Pearson's R^2 = 0.2591, 0.3176, 0.1357 and 0.1376 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively, all $p < 0.05$). Moreover, across the surface of the RS filter bed the microbial community structure changed more than the microbial community membership (slope of the regression line = 1.09 x 10^{-2}, 9.48 x 10^{-3}, 2.26 x 10^{-3} and 2.27 x 10^{-3} for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively). In particular, from the structure-based plots, the estimated change in the microbial community structure between the two most distant separated points were approximately between 9% and 12%, depending on the structure based dissimilarity measure used (d_{BC}, M ± SD = 0.25 ± 0.10 and 0.37 ± 0.15 for 5 m and 15 m, respectively; d_{WUF}, M ± SD = 0.16 ± 0.08 and 0.25 ± 0.10 for 5 m and 15 m, respectively) (Figures 3.14a and 3.14b); whereas the estimated change in the microbial community membership between the two most distant separated points were approximately between 1% and 2%, depending on the membership-based dissimilarity measure used (d_J, M ± SD = 0.69 ± 0.04 and 0.70 ± 0.06 for 5 m and 15 m, respectively; d_{UUF}, M ± SD = 0.63 ± 0.05 and 0.65 ± 0.06 for 5 m and 15 m, respectively) (Figures 3.14c and 3.14b) (see other comparisons Table 3.3).

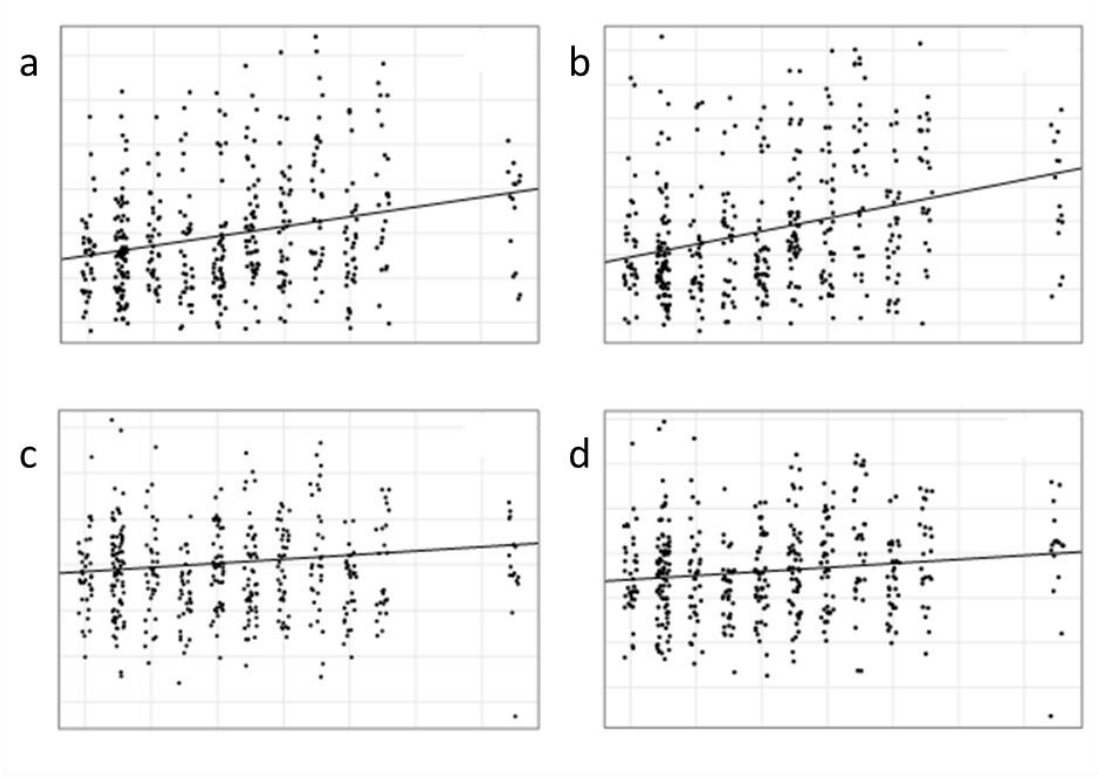


Figure 3.14 Correlations between (a) Bray-Curtis, (b) weighted UniFrac, (c) Jaccard and (d) unweighted UniFrac distances of microbial communities sampled at any two surface locations situated across the surface of the rapid sand filter bed. See Table 3.3 for means and standard deviations.

Table 3.3 Means and standard deviations of structure and membership-base distances between sites across the surface of the rapid sand filter bed

	Structure-based diversity matrices				**Membership-based diversity matrices**			
Distance	**Bray-Curtis**		**Weighted UniFrac**		**Jaccard**		**Unweighted UniFrac**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
2 m	0.25	0.10	0.16	0.08	0.69	0.04	0.63	0.05
3 m	0.26	0.11	0.15	0.08	0.69	0.05	0.63	0.05
4 m	0.28	0.10	0.16	0.09	0.69	0.05	0.62	0.06
5 m	0.27	0.14	0.17	0.08	0.67	0.04	0.61	0.04
6 m	0.26	0.13	0.16	0.08	0.70	0.04	0.62	0.05
7 m	0.32	0.13	0.20	0.09	0.69	0.05	0.63	0.05
8 m	0.31	0.13	0.21	0.10	0.70	0.04	0.64	0.05
9 m	0.40	0.16	0.26	0.12	0.71	0.07	0.66	0.06
10 m	0.28	0.12	0.18	0.08	0.68	0.04	0.62	0.04
11 m	0.39	0.16	0.26	0.10	0.70	0.05	0.65	0.05
15 m	0.37	0.15	0.25	0.10	0.70	0.06	0.65	0.06

Abbreviations: ***M***, mean; ***SD***, standard deviation.

Similarly, as shown in Figure 3.15 significant positive Pearson's correlations between distances across the depth of the RS filter bed and differences in the microbial community memberships and structures were found (Pearson's R^2 = 0.2255, 0.2194, 0.4185 and 0.3907 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively, all $p < 0.05$). Moreover, along the depth of the RS filter bed the microbial community membership changed more than the microbial community structure (slope of the regression line = 3.98E-03, 3.79E-03, 3.63E-03 and 3.62E-03 for d_{BC}, d_{WUF}, d_J and d_{UUF}, respectively). In particular, from the membership-based plots, the estimated change in the microbial community membership between the two most distant separated points were approximately between 10% and 11%, depending on the membership-based dissimilarity measure used (d_J, M ± SD = 0.70 ± 0.05 and 0.81 ± 0.07 for 5 cm and 30 cm, respectively; d_{UUF}, M ± SD = 0.64 ± 0.05 and 0.74 ± 0.08 for 5 cm and 30 cm, respectively) (Figures 3.15c and 3.15d); whereas the estimated change in the microbial community structure between the two most distant separated points were approximately 10%, depending on the structure based dissimilarity measure used (d_{BC}, M ± SD = 0.28 ± 0.11 and 0.21 ± 0.09 for 5 cm and 30 cm, respectively; d_{WUF}, M ± SD = 0.38 ± 0.13 and 0.31 ± 0.14 for 5 cm and 30 cm, respectively) (Figures 3.15a and 3.15b) (see other comparisons Table 3.4).

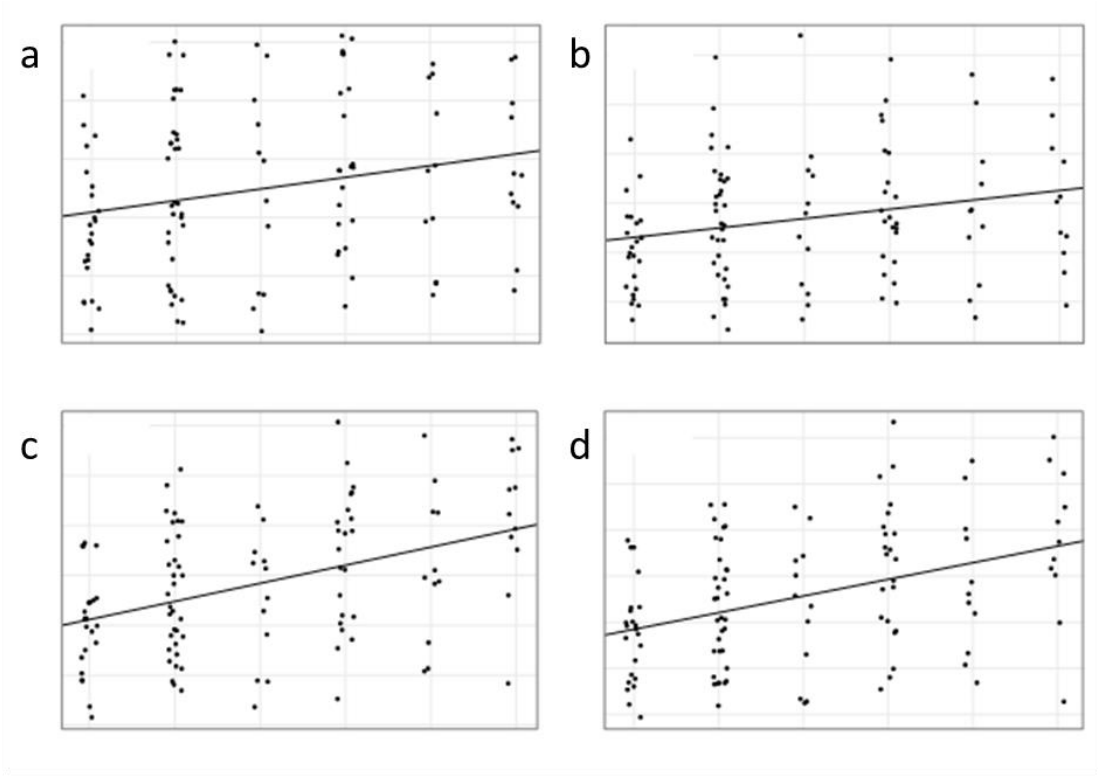


Figure 3.15 Correlations between (a) Bray-Curtis, (b) weighted UniFrac, (c) Jaccard and (d) unweighted UniFrac distances of microbial communities sampled at any two core locations situated along the depth of the rapid sand filter bed. See Table 3.4 for means and standard deviations.

Table 3.4 Means and standard deviations of structure and membership-base distances between sites along the depth of the rapid sand filter bed

	Structure-based diversity matrices				Membership-based diversity matrices			
Distance	**Bray-Curtis**		**Weighted UniFrac**		**Jaccard**		**Unweighted UniFrac**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
5 cm	0.28	0.11	0.21	0.09	0.70	0.05	0.64	0.05
10 cm	0.35	0.14	0.27	0.13	0.73	0.06	0.67	0.06
15 cm	0.35	0.17	0.27	0.16	0.73	0.07	0.67	0.08
20 cm	0.40	0.14	0.30	0.13	0.77	0.07	0.71	0.08
25 cm	0.37	0.15	0.29	0.16	0.76	0.08	0.69	0.08
30 cm	0.38	0.13	0.31	0.14	0.81	0.07	0.74	0.08

Abbreviations: ***M***, mean; ***SD***, standard deviation.

3.3.2 Drinking Water Treatment Plant 2

Sequence data

The Illumina MiSeq platform generated a total number of 3 643 232 sequence reads from 90 Rapid Sand (RS) filter bed sand samples (including both surface and depth samples) collected at the second treatment plant. As a result of failed sequences runs, four samples (Sand_35, Sand_78, Sand_91 and Sand_94) had to be removed from the original database prior to processing. After the series of quality processing steps as described above, 2 623 152 sequence reads were retained in the final database (Table A20). This number of sequence reads generated 15 097 Operational Taxonomic Units (OTUs). For the purpose of the alpha diversity analysis, all samples were sub-sampled a 1 000 times to have 4 648 sequence reads in each sample (the minimum number of sequences in a sample). This step ensured that downstream diversity analyses compared samples with the same sequencing depth. In addition, for beta diversity analysis purposes, the original database was filtered/truncated by removing OTUs which (i) had an abundance below 1% of the total number of sequences in the original dataset and (ii) was not present in at least 75% of the 90 samples. This new beta diversity database (or the subset database) consisted of 89% of the total sequence reads (n = 2 346 217), which resulted in keeping 1.66% of the total number of OTUs identified (n = 251). Based on a Mantel's test (r = 0.8667, p = 0.001), this subset of OTUs (n = 251) still explained 87% of variation within the microbial community of the RSF.

Community diversity

The alpha diversity metrics identified the richness/observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J) of each RS filter surface and depth sample. An analysis of variance (ANOVA) was applied in order to determine whether the variation between the S_{obs}, H' and J of each samples was significant (p-value < 0.05) or not (p value > 0.05). Despite subsampling, most of the species identified were still represented in the RSF database as the Good's coverage was in the range

of 93-99%. Based on the findings there was a significant difference in the alpha diversity of the microbial community inhabiting RS filter surface samples (ANOVA p-value < 0.05 for all alpha diversity metrics) (Table 3.5a and Table 3.6a) based on temporal groupings. The number of species observed (S_{obs}) showed a change, moving from one month to the other. An increase in the observed species from the summer (M ± SD = 360.18 ± 95.73) to autumn (ranging from M ± SD = 402.82 ± 120.81 to M ± SD = 416.62 ± -46.43) months was evident, followed by a decrease from the autumn to the winter (M ± SD = 324.05 ± 99.88 to M ± SD = 399.73 ± 120.91) months. Fluctuations in the diversity (H') were observed irrespective of the month but in contrast, the evenness of the microbial community stayed relatively constant (Table 3.7). Investigating the microbial community within the RS filter surface samples located in a single filter, showed that the observed species (S_{obs}) and their diversity (H') changed frequently according to seasonal fluctuations (Table 3.5b, Tukey's post-hoc HSD test, p <0.05). The only change in the microbial community evenness occurred between the autumn months (March, April and May) and the first winter month (June) (Tukey's post-hoc HSD test, p < 0.05, Table 3.5b). The microbial community harboured in the RS filter surface samples, located in different filters, also supported the observation that temporal changes influence the diversity (H'), evenness (J) and especially the richness (S_{obs}) in the microbial community (Tukey's post-hoc HSD test, p < 0.05, Table 3.6b). Interestingly, no significant temporal changes occurred in the alpha diversity of the microbial community identified in the RS filter depth samples (ANOVA, p > 0.05 for all alpha diversity metrics).

Table 3.5a Analysis of Variance (ANOVA) testing the significant difference of the richness, diversity and evenness of the microbial community in RS filter on a monthly basis from the surface samples located in a single filter

Observed species (S_{obs})	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	126310	25261.9	12.22	1.303×10^{-7}
Within groups	47	97163	2067.3		
Total	52	223473			
Shannon index (H')	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	1.71	0.34	12.14	1.41×10^{-7}
Within groups	47	1.33	0.03		
Total	52	3.04			
Pielou's evenness (J)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	0.02	3.23×10^{-3}	7.21	4.46×10^{-5}
Within groups	47	0.02	4.48×10^{-4}		
Total	52	0.04			

Abbreviations: degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic (FST).

Table 3.5b Tukey's post-hoc HSD test, showing between which months the significant difference in alpha-diversity in the microbial community in the RS filter surface samples located in a single filter occurs

	Alpha-diversity metrics		
	Observed species (S_{obs})	**Shannon index (H')**	**Pielou's evenness (J)**
February-April	3.06×10^{-4}	2.71×10^{-2}	-
June-April	6.00×10^{-6}	1.30×10^{-5}	2.74×10^{-3}
July-February	9.43×10^{-3}	-	-
March-February	3.73×10^{-3}	3.86×10^{-3}	-
May-February	2.29×10^{-3}	2.09×10^{-2}	-
June-July	2.32×10^{-4}	1.01×10^{-3}	-
March-June	8.06×10^{-5}	9×10^{-7}	3.12×10^{-5}
May-June	4.69×10^{-5}	7.2×10^{-6}	5.74×10^{-4}

Table 3.6a Analysis of Variance (ANOVA) testing the significant difference of the richness, diversity and evenness of the microbial community in RS filter on a monthly basis from the surface samples located in different filters

Observed species (S_{obs})	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	40675	8135	8.83	1.03×10^{-3}
Within groups	12	11055	921.25		
Total	17	51730			
Shannon index (H')	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	0.66	0.13	7.29	2.37×10^{-3}
Within groups	12	0.22	0.02		
Total	17	0.88			
Pielou's evenness (J)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	5	7.26×10^{-3}	1.45×10^{-3}	5.72	6.32×10^{-3}
Within groups	12	3.05×10^{-3}	2.54×10^{-4}		
Total	17	0.01			

Table 3.6b Tukey's post-hoc HSD test, showing between which months the significant difference in alpha-diversity in the microbial community in the RS filter surface samples located in different filter occurs

	Alpha-diversity metrics		
	Observed species (S_{obs})	**Shannon index (H')**	**Pielou's evenness (J)**
February-April	5.87×10^{-2}	-	-
June-April	7.90×10^{-3}	1.52×10^{-2}	4.03×10^{-2}
March-February	1.12×10^{-2}	-	-
March-July	-	3.67×10^{-2}	4.66×10^{-2}
March-June	1.61×10^{-3}	2.14×10^{-3}	5.42×10^{-3}
May-June	1.48×10^{-2}	2.02×10^{-2}	4.56×10^{-2}

Abbreviations: degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic (FST).

Table 3.7 The average (M ± SD) alpha diversity (richness (S_{obs}), diversity (H') and evenness (J)) of the microbial community in RS filter across all the months within the six-month sampling period

Month	**Observed species (S_{obs})**	**Shannon index (H')**	**Pielou's evenness (J)**
February	360.18 ± 95.73	3.90 ± 0.65	0.66 ± 0.09
March	413.85 ± 120.81	4.20 ± 0.55	0.70 ± 0.06
April	402.82 ± 146.43	3.96 ± 0.83	0.66 ± 0.1
May	416.62 ± 127.8	4.06 ± 0.76	0.68 ± 0.1
June	324.05 ± 99.88	3.69 ± 0.69	0.64 ± 0.1
July	399.73 ± 120.91	4.07 ± 0.39	0.68 ± 0.04

For the beta diversity analyses, a subset of the RS filter community was analysed to minimise skewing of data due to rare species. Based on a Mantel's test ($r = 0.8508$, $p = 0.001$), this subset of OTUs ($n = 220$) still explained 85% of changes that could occur in the microbial community structure and membership of the RS filter samples. The beta diversity metrics used to analyse the community structure was Bray-Curtis dissimilarity metric and weighted UniFrac metric and community membership were analysed by performing the Jaccard dissimilarity metric and unweighted UniFrac. Along with these metrics, an Analysis of Molecular Variance (AMOVA) was performed using structure-based and membership-based dissimilarity measures to test for significant differences between spatial and temporal groupings

As observed with the alpha diversity metrics, the beta diversity metrics also indicated that the changes identified in the RS filter microbial community was linked to temporal changes (AMOVA $p < 0.05$ for all beta diversity metrics). A positive correlation was observed between the increased difference between consecutive months and an increase in dissimilarity, based on a temporal decay relationship (Figure 3.16). Focusing specifically on community membership (Jaccard metric), it was noticed that half of the comparisons were non-significant ($p > 0.05$). These pairwise comparisons included the months April and May (autumn months) as well as June and July (winter months) (Table 3.8).

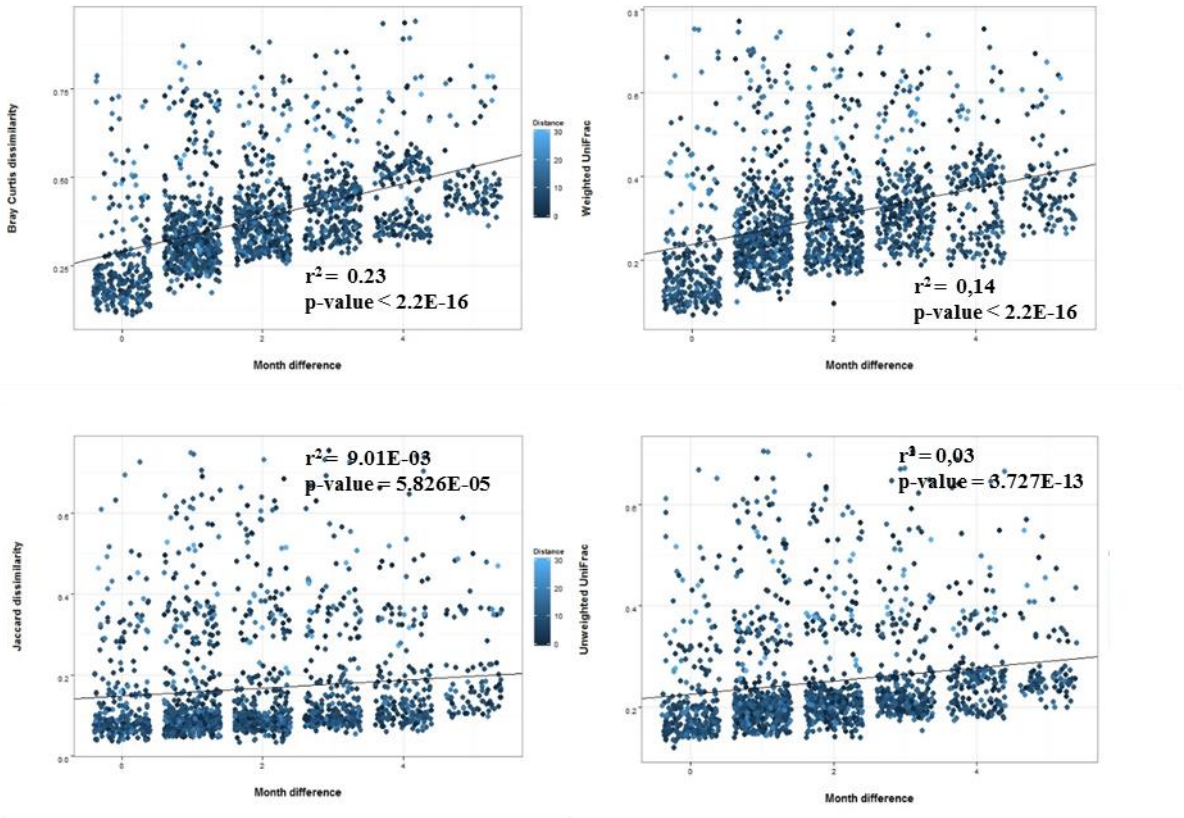


Figure 3.16 Temporal decay relationship illustrating how temporal trends can influence the RS filter community structure (A) Bray-Curtis dissimilarity and (B)Weighted UniFrac and community membership (C) Jaccard dissimilarity and (D) Unweighted UniFrac as a based on month difference

Table 3.8 Analysis of Molecular Variance (AMOVA) showing the temporal changes in the microbial community membership of RS filter using the Jaccard dissimilarity pairwise metric.

Jaccard dissimilarity metric					
April vs February	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 1	0. 1	2. 02	0. 06
Within	28	1. 36	0. 05		
Total	29	1. 45			
April vs July	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 08	0. 08	1. 85	0. 11
Within	26	1. 08	0. 04		
Total	27	1. 15			
April vs June	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 06	0. 06	1. 16	0. 26
Within	26	1. 46	0. 06		
Total	27	1. 52			
April vs March	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 04	0. 04	0. 97	0. 32
Within	28	1. 17	0. 04		
Total	29	1. 21			
April vs May	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 05	0. 05	1. 36	0. 23
Within	28	1. 06	0. 04		
Total	29	1. 11			
February vs July	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 14	0. 14	4. 64	1.00×10^{-3}
Within	28	0. 84	0. 03		
Total	29	0. 98			
February vs June	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 13	0. 13	2. 94	1.00×10^{-3}
Within	28	1. 22	0. 04		
Total	29	1. 35			
February vs March	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 09	0. 09	2. 93	4.00×10^{-3}
Within	30	0. 9	0. 03		
Total	31	1. 03			
February vs May	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 11	0. 11	4. 12	1.00×10^{-3}
Within	30	0. 82	0. 03		
Total	31	0. 93			

July vs June	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 03	0. 03	0. 81	0. 54
Within	26	0. 94	0. 04		
Total	27	0. 97			
July vs March	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 06	0. 06	2. 62	2. 00 x 10^{-3}
Within	28	0. 66	0. 02		
Total	29	0. 72			
July vs May	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 03	0. 03	1. 75	0. 11
Within	28	0. 54	0. 02		
Total	29	0. 57			
June vs March	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 06	0. 06	1. 68	0. 11
Within	28	1. 03	0. 04		
Total	29	1. 1			
June vs May	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 05	0. 05	1. 41	0. 14
Within	28	0. 92	0. 03		
Total	29	0. 96			
March vs May	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0. 03	0. 03	1. 37	0. 2
Within	30	0. 63	0. 02		
Total	31	0. 66			

Abbreviations: degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic (FST).

Community membership

As already described, the subset of the data was attained by filtering the original RS filter dataset and removing all OTUs which had an abundance below 1% of the total number of sequences in the original dataset and did not occur in more than 75% of the 90 RS filter samples. A relatively clear picture of the "who" is in the RS filter was observed from this subset database. The subset of data contained 2 346 217 sequences, which were identified and classified at 97% similarity cut-off into 251 OTUs. Based on the abundance of the 251 OTUs, it is clear that the microbial community in the RS filter is homogenous (Figure 3.17).

Focusing on the 251 OTUs, it was determined that the majority of the microbial community samples dominated by bacteria, which had a *MRA* of 98.29 ± 0.99% across all RS filter samples. Archaea was present with a MRA of 1.71 ± 0.99%. Within this subset microbial community 19 phyla were identified, where only one phylum, *Crenarchaeota* (MRA = 1.72 ± 0.99%), was affiliated to the Archaea kingdom.

Although the bacteria in the microbial community in the RS filter samples were quite diverse, *Proteobacteria* dominated the microbial community with a MRA of 41.77 ± 19.03% across all RS filter samples. This phylum was followed by *Actinobacteria* (MRA = 20.30 ± 7.73%), *Bacteroidetes* (MRA = 16.65 ± 6.87%), *Cyanobacteria* (MRA = 5.35 ± 3.35%) and *Acidobacteria* (MRA = 5.32 ± 2.40%), respectively (Figure 3.18).

Of the 251 OTUs, the most abundant OTU (OTU1) was classified to the *Actinobacteria* phylum, *Actinobacteria* class and was unclassified at the genus level with a MRA of 12.81 ± 5.60%. Both OTU2 (MRA = 10.41 ± 16.02%) and OTU3 (MRA = 5.83 ± 5.11%) were affiliated to the *Proteobacteria* phylum and the *Gamma-* and *Betaproteobacteria* classes, respectively. At the genus level OTU2 was classified as *Pseudomonas* and OTU3 was classified as *Rhodoferax.* The two remaining OTUs, OTU4 and OTU5, where classified as *Bacteroidetes* (MRA = 4.00 ± 4.60%) and *Cyanobacteria* (MRA = 5.02 ± 3.24%) at the phylum level, respectively. OTU4 was identified as *Flavobacterium* (class, *Flavobacteria*) and OTU5 as *Synechococcus* (class, *Synechococcophycideae*).

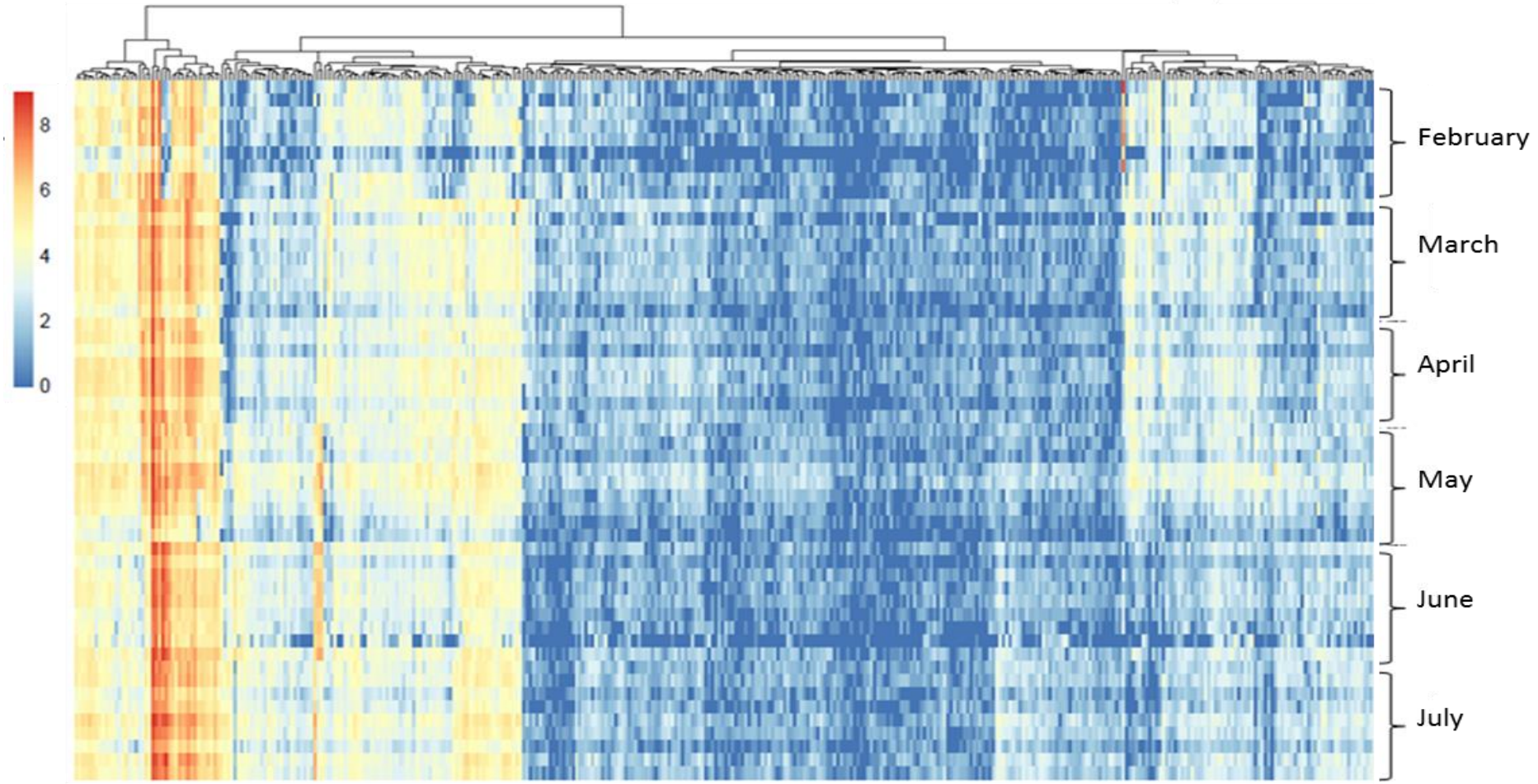


Figure 3.17 Heatmap constructed using the absolute sequence abundances (log transformed) of 251 OTUs that were selected based on their relative abundance within a sample (i.e. relative abundance threshold ≥ 1%) and by the percentage of samples in which they were detected (i.e. detection frequency threshold ≥ 75%). The heatmap boxes were colored from red-to-blue to represent higher-to-lower abundances, and OTUs not represented by a sequence were assigned 10^{-6} (displayed as dark blue).

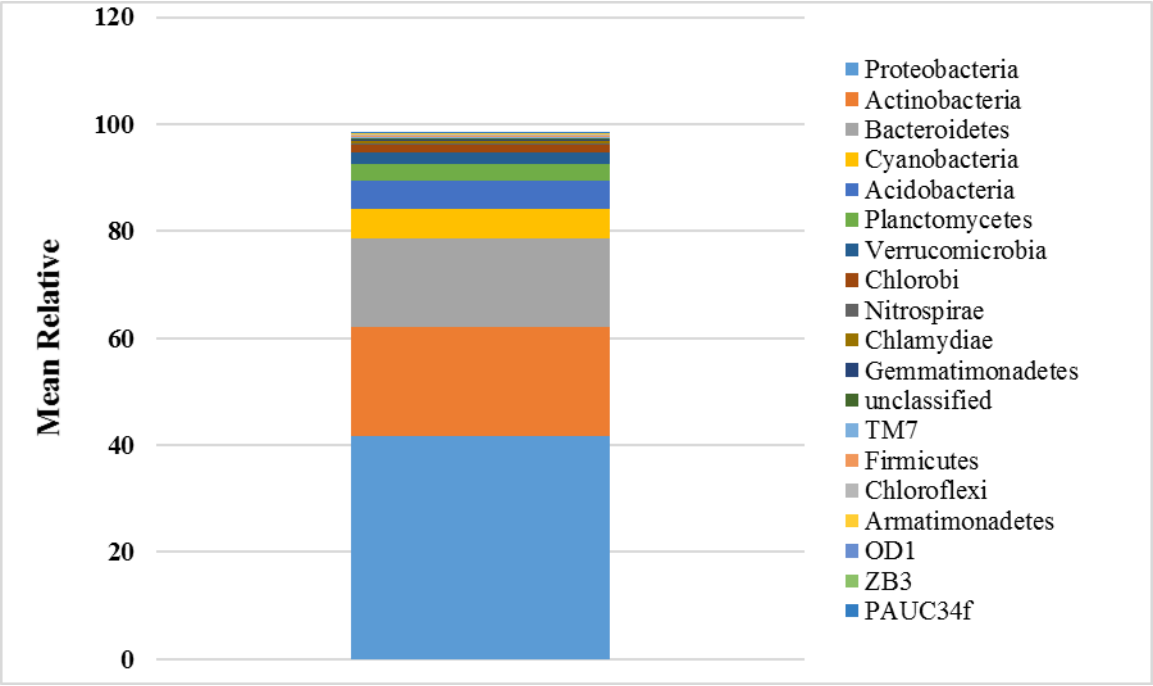


Figure 3.18 Stacked bar charts presenting the mean relative abundance of the bacterial phyla identified and classified in RS filter sand samples.

For the other classification levels, only the three most dominant phyla were focused on. Within these three dominant phyla, 14 bacterial classes were present. Of these 14 bacterial classes, the five most dominant classes across RSF sand samples were in descending order i.e. *Gammaproteobacteria* (phylum *Proteobacteria;* MRA = 18.35 ± 18.07%), *Actinobacteria* (phylum *Actinobacteria;* MRA = 17.56 ± 7.03%), *Betaproteobacteria* (phylum *Proteobacteria;* MRA = 12.36 ± 5.17%), *Alphaproteobacteria* (phylum *Proteobacteria;* MRA = 10.94 ± 6.34%) and *Flavobacteria* (phylum *Bacteroidetes;* MRA = 7.33 ± 4.72%) (Figure 3.19). Within the three most dominant phyla 30 orders were identified. The order *Actinomycetales* (*Actinobacterium*) was the most dominant amongst all the 30 orders, with a MRA of 17.56 ± 7.03%. *Flavobacteriales* (MRA = 7.33 ± 4.72%) represented the phylum *Bacteroidetes*. Amongst the class *Gammaproteobacteria* the classes *Pseudomonadales* (MRA = 12.14 ± 17.10%) and *Aeromonadales* (MRA = 2.88 ± 10.31%) had the highest MRA. For the *Betaproteobacteria* the MRA of the order *Burkholderiales* was 9.68 ± 5.10% and members belonging to unclassified order had a MRA = 1.60 ± 0.70%. Finally, in the class *Alphaproteobacteria*, *Sphingomonadales* (MRA = 4.78 ± 3.36%) and *Rhodobacterales* (MRA = 1.90 ± 0.97%) were the most abundant orders.

Within the subset of data, 40 genera were present. The three genera that dominated were the unclassified genera from the *Actinobacteria* phylum, *Pseudomonas* (*Proteobacteria* phylum) and the unclassified genera from the *Bacteroidetes* phylum, with MRA of 19.77 ± 7.62%, 11.68 ± 16.61% and 8.45 ± 4.68%, respectively. The genera with the two highest MRA in the *Gammaproteobacteria* class were *Pseudomonas* (MRA = 11.68 ± 16.61%, *Pseudomonadales* (Order)) and *Aeromonas* (MRA = 2.88 ± 10.31%, *Aeromonadales* (Order)). The

genus *Rhodoferax* (MRA = 5.83 ± 5.12%, *Burkholderiales* (Order)) and other unclassified genera (MRA = 4.54 ± 1.78%) were the most abundant *Betaproteobacteria* in the subset data. The *Alphaproteobacteria* were represented by the unclassified genera (MRA = 3.42 ± 3.75%) and *Sphingomonas* (MRA = 2.45 ± 3.27%, *Sphingomonadales* (Order)).

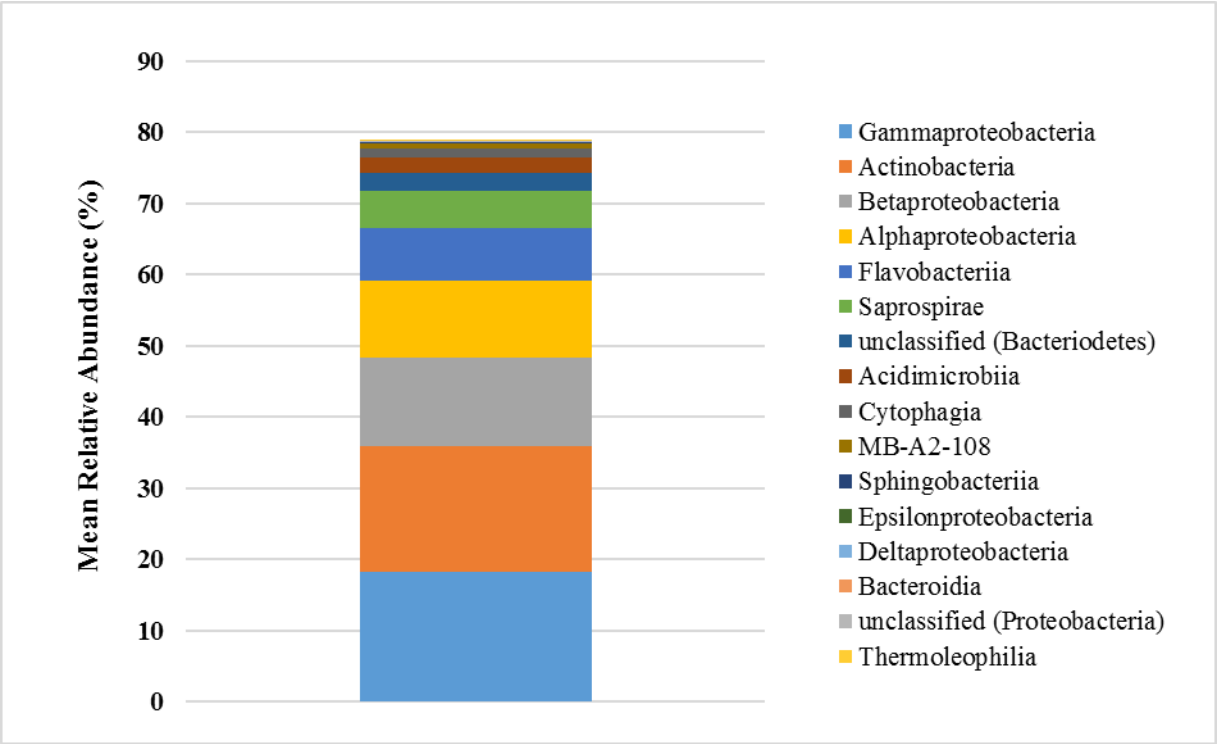


Figure 3.19 Stacked bar charts presenting the mean relative abundance of the bacterial classes identified and classified in RS filter samples, which are associated with the three most dominant phyla.

Spatial trends amongst rapid sand filter communities

The same metrics (S_{obs}, H' and J) and significance test (ANOVA) that were applied to investigate the possible temporal variation in the RS filter microbial community reported above, were used to investigate whether the variation within the RS filter microbiome can be explained by spatial groupings, i.e. surface and/or depth. However, there were no significant differences observed (ANOVA, $p > 0.05$ for all metrics) in the mean richness (S_{obs}), diversity (H') and evenness (J) (i) across the surface of a single RS filter bed (S_{obs}, $F_{ST} = 0.35$, H', $F_{ST} = 0.47$ and J, $F_{ST} = 0.52$), (ii) along the depth of the RS filter bed (S_{obs}, $F_{ST} = 0.38$, H', $F_{ST} = 0.44$ and J, $F_{ST} = 0.44$) and (iii) across the surface samples between different RS filter beds (S_{obs}, $F_{ST} = 0.17$, H', $F_{ST} = 0.08$ and J, $F_{ST} = 0.19$). Even though there was no significant difference in the overall RSF microbial community, some spatial variations were observed.

The slight spatial variation occurred between the richness (S_{obs}) and diversity (H') of all RS filter samples whereas the evenness (J) of the observed species within all the RS filter samples remained constant (Table 3.9, 3.10 and 3.11). In terms of the surface samples, the only changes observed, were in the richness (S_{obs}) of

the RS filter samples. Between the samples from the same filter bed, fluctuations are evident, with a difference between the samples, which had the maximum (Surface_1, S_{obs} M ± SD = 481.08 ± 68.26) and minimum (Surface_8, S_{obs} M ± SD = 432.88 ± 98.76) number of species. For the microbial communities associated with the surface samples from different filter beds, only one filter bed (Surface_10, S_{obs} M ± SD = 458.37 ± 55.20) showed a decrease in the richness whereas the other two filter beds, Surface_5 (S_{obs} = M ± SD 471.31 ± 57.53) and Surface_11 (S_{obs} M ± SD = 477.72 ± 61.29) were similar. Concerning the depth samples, a clear decrease in the richness and diversity was observed between the top layers of the RS filter, Depth 1 and Depth 2 (Depth_1, S_{obs} M ± SD = 382.54 ± 167.75, H' M ± SD = 3.79 ± 1.06 and Depth_2, S_{obs} M ± SD = 336.76 ± 125.19, H' M ± SD = 3.53 ± 1.22) and deepest layers Depth 4 and Depth 5 (Depth_4, S_{obs} M ± SD = 302.46 ± 172.11, H' M ± SD = 3.43 ± 1.34 and Depth_5, S_{obs} M ± SD = 285.52 ± 99.03, H' M ± SD = 3.15 ± 0.69).

Table 3.9 The average (M ± SD) alpha diversity data (richness, diversity and evenness) of each RS filter surface sample site in a single filter bed

Location	Observed species (S_{obs})	Shannon index (H')	Pielou's evenness (J)
Surface_1	481.08 ± 68.26	4.30 ± 0.33	0.70 ± 0.04
Surface_2	457.89 ± 56.73	4.34 ± 0.21	0.71 ± 0.03
Surface_3	469.32 ± 43.31	4.34 ± 0.24	0.71 ± 0.03
Surface_4	477.27 ± 41.83	4.41 ± 0.20	0.71 ± 0.02
Surface_5	471.31 ± 57.53	4.36 ± 0.25	0.71 ± 0.03
Surface_6	439.94 ± 103.66	4.28 ± 0.27	0.71 ± 0.03
Surface_7	459.30 ± 59.94	4.27 ± 0.24	0.70 ± 0.03
Surface_8	432.88 ± 98.76	4.16 ± 0.30	0.69 ± 0.03
Surface_9	473.58 ± 59.16	4.33 ± 0.19	0.70 ± 0.02

Table 3.10 The average (M ± SD) alpha diversity data (richness, diversity and evenness) of each RS filter depth sample site in a single filter bed

Location	Observed species (S_{obs})	Shannon index (H')	Pielou's evenness (J)
Depth_1	382.54 ± 167.75	3.79 ± 1.06	0.64 ± 0.13
Depth_2	336.76 ± 125.19	3.53 ± 1.22	0.60 ± 0.18
Depth_3	258.28 ± 148.97	3.05 ± 1.04	0.55 ± 0.14
Depth_4	302.46 ± 172.11	3.43 ± 1.34	0.60 ± 0.18
Depth_5	285.52 ± 99.03	3.15 ± 0.69	0.56 ± 0.09

Table 3.11 The average (M ± SD) alpha diversity data (richness, diversity and evenness) of each RS filter surface sample site in different filter beds

Location	Observed species (S_{obs})	Shannon index (H')	Pielou's evenness (J)
Filter bed 1 (Surface_5)	471.31 ± 57.53	4.36 ± 0.25	0.71 ± 0.03
Filter bed 2 (Surface_10)	458.37 ± 55.20	4.32 ± 0.20	0.70 ± 0.02
Filter bed 3 (Surface_11)	477.72 ± 61.29	4.32 ± 0.27	0.70 ± 0.03

The same beta diversity metrics (i.e. Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac) and subset of OTUs as descried above, were used to investigate whether a spatial trend was evident amongst the microbial communities associated with the RS filter. The beta diversity metrics clearly displayed that there were no significance differences in the microbial community structure and membership (i) across the surface of a single RS filter bed, (ii) along the depth of the same RS filter bed and (iii) between different RS filter beds (AMOVA $p > 0.05$ for all beta diversity metrics). The beta diversity metrics also indicated that the community structure and community membership of the RS filter surface samples of the same filter bed, those of different filter beds and depth samples were very similar, with an average Bray-Curtis dissimilarity range of M ± SD = 0.34-0.5 ± 0.09-0.16 (min = 0.09-0.13, max = 0.5-0.81), average Jaccard dissimilarity ranges of M ± SD = 0.08-0.58 ± 0.02-0.16 (min = 0.03-0.17, max = 0.14-0.94), average weighted UniFrac ranges of M ± SD = 0.24-0.36 ± 0.08-0.17 (min = 0.04-0.05, max = 0.45-0.73) and average unweighted UniFrac ranges of M ± SD = 0.04-0.29 ± 0.02-0.17 (min = 0, max = 0.1-0.65).

A Pearson's correlation test was performed, indicating that the surface samples from the same filter bed had a slightly negative correlation with spatial groupings (Pearson's R for Bray-Curtis dissimilarity = -0.06, Pearson's R for Jaccard dissimilarity = -0.10, Pearson's R for weighted UniFrac = -0.07 and Pearson's R for unweighted UniFrac = -0.11). In contrast, the RS filter surface samples from different filter beds, resulted in slight positive correlations with spatial groupings (Pearson's R for Bray-Curtis dissimilarity = 0.090, Pearson's R for Jaccard dissimilarity = 0.11, Pearson's R for Weighted UniFrac = 0.08 and Pearson's R for Unweighted UniFrac = 0.13).

In contrast to the previous system evaluated, no significant differences were observed between samples taken on the surface of the filter and those taken deeper within the filter bed (core samples). Even when the distance decay along the depth of the RS filter was analysed, no significant differences were observed for both the community structure and community membership (Figure 3.17).

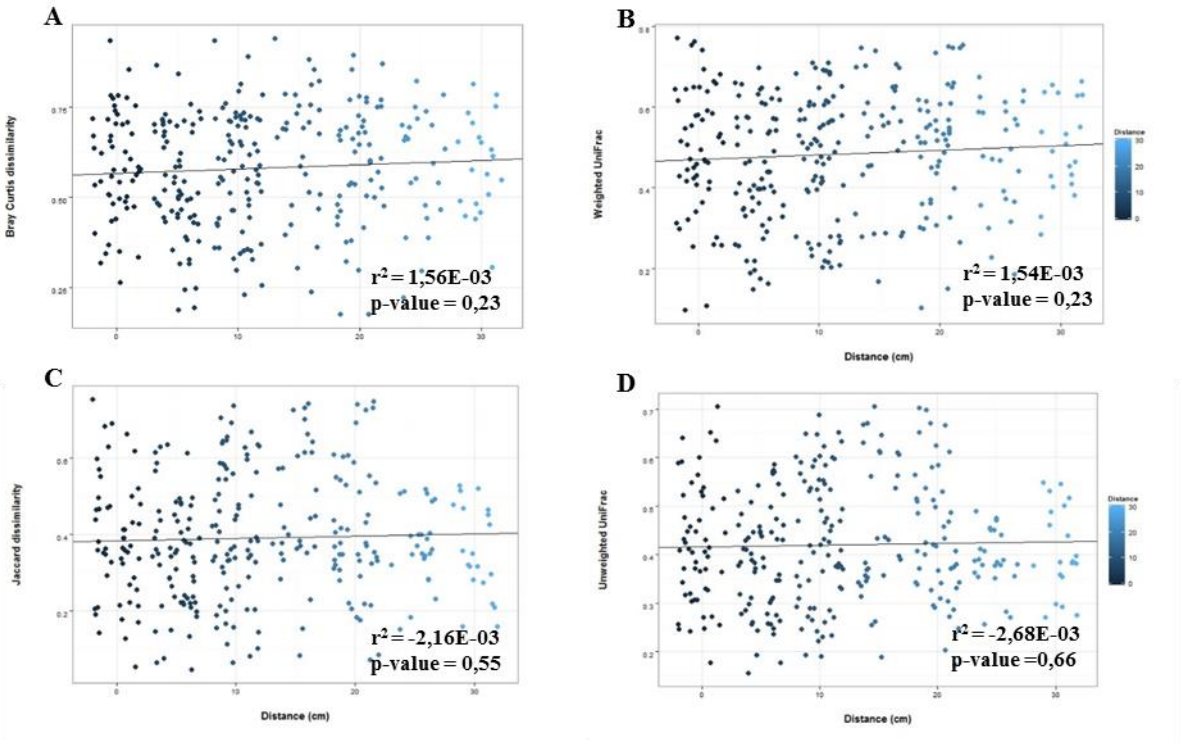


Figure 3.20 Distance-decay relationship along the depth of the filter bed based on microbial community structure, (A) Bray-Curtis pairwise distance metric and (B) Weighted UniFrac metrics, and the microbial community membership (C) Jaccard pairwise distance metric and (D) Unweighted UniFrac.

3.4 DISCUSSION

3.4.1 Community membership

An initial 4-month survey in a large scale DWTP was conducted to assess the extent of spatial and temporal variation amongst RS filter microbial communities, by analysing 188 samples through V4 16S rRNA profiling. This was followed by a second study where three rapid sand filters were sampled over a period of 6 months within another large DWTP. For this study 90 samples were analysed. A disadvantage regarding the methodology is that a method to discriminate between viable and non-viable cells was not included. Here, the OTUs were classified from the processed V4 16S rRNA gene sequences that were amplified from the total DNA extracts of the samples that potentially contains DNA of dead or damaged cells. Researchers have sounded a note of caution with regards to this approach and have proposed several methods to circumvent for this bias, i.e. membrane damage (Nocker *et al*., 2007; Chiao *et al*., 2014), RNA (Eichler *et al*., 2006; Keinänen-Toivola *et al*., 2006), enzymatic activity (Hoefel *et al*., 2005) and DNA damage (McCarty and Atlas, 1993). However, as stated previously, none of these methods, alone or in combination, have yet provided robust discrimination between viable and non-viable cells (Pinto *et al*., 2014). Therefore, characterisation of the microbial communities in this study encompasses the description of both viable and non-viable microbial species, using a universal approach that is similar to previous studies reported in literature (Kwon *et al*., 2011;

Pinto *et al*., 2012; Kim *et al*., 2014; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Albers *et al*., 2015; LaPara *et al*., 2015; Gülay *et al*., 2016).

As deduced from the rarefaction curves obtained from samples collected from DWTP 1, the classified OTUs did not encompass all OTUs, which implies that the microbial communities were more diverse than explained and that further or deeper sequencing may be required to increase the number of detected OTUs. However, as documented previously, deeper sequencing will not have a significant impact on the structure of the microbial communities, as it will likely only improve the representation of low-abundant microbial species or OTUs (Pinto and Raskin, 2012). Moreover, the large amount of sequences obtained in this study, compared to those generated using former 16S rRNA-based culture-independent fingerprinting techniques (i.e. denaturing gel electrophoresis (DGGE) and terminal-restriction fragment length polymorphism (T-RFLP)) combined with sequencing (Fonseca *et al*., 2001; Kasuga *et al*., 2007; Feng *et al*., 2012; Liao *et al*., 2012; Feng *et al*., 2013) still provides a significantly more detailed and accurate description of the microbial communities present. Therefore, it is plausible to assume that a nearly, complete and accurate description of the microbial community structure was given in this study.

During the first study (DWTP 1) 5 708 microbial OTUs were identified. These OTUs were distributed across 61 different phyla, of which only *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* were regarded as dominant as they constituted for approximately 95% of the microbial community. These phyla are commonly found freshwater systems (i.e. rivers, lakes and dams) (Newton *et al*., 2011; Martinez-Garcia *et al*., 2012; Llirós *et al*., 2014) and groundwater systems (i.e. aquifers) (Gülay *et al*., 2016; Albers *et al*., 2015) that continuously feed DWTPs, and are often amongst the most dominant phyla reported in rapid gravity sand (RGS) filtration systems (i.e. RS filters and granular activated carbon (GAC) filters) (Lin *et al*., 2014). In congruence with previous studies, *Proteobacteria* was the most dominant phylum, and constituted for approximately 86.44% of the RS filter microbial community (Pinto *et al*., 2012; Bai *et al*., 2013; Kim *et al*., 2014; Lautenschlager *et al*., 2014; Lin *et al*., 2014; Albers *et al*., 2015; Gülay *et al*., 2016; Li *et al*., 2017).

Furthermore, within *Proteobacteria*, *Gammaproteobacteria* was the most common class in all RS filter samples, and accounted for approximately 90.25% of the *Proteobacteria* sequences. In general, *Gammaproteobacteria* are more abundant in saltwater systems (i.e. oceans (Biers *et al*., 2009) and saline lakes (Wu *et al*., 2006)) than in freshwater systems (Newton *et al*., 2011). However, *Gammaproteobacteria* have been reported to be amongst the most dominant proteobacterial classes in RGS filter systems of large-scale DWTPs located in Copenhagen, Denmark (Gülay *et al*., 2016; Albers *et al*., 2015) and Hubei, China (Lin *et al*., 2014). Previously, Zavarzin *et al*. (1991) suggested that species of *Gammaproteobacteria* are copiotrophs, which are defined as species that thrive in environments rich in organic matter (Koch, 2001). Generally, RGS filtration systems can be rich in organic matter as they are designed to facilitate the removal of natural (or total) organic matter (NOM or TOM) constituents, including dissolved organic matter (i.e. dissolved organic carbon (DOC), dissolved organic nitrogen (DON) and dissolved organic phosphorous (DOP)) and undissolved organic matter (i.e. particulate organic carbon (POC)) (Pagano *et al*., 2014), to deliver low-nutrient oligotrophic filtered water, which reduces the growth potential of microbial species during

drinking water distribution (Liu *et al*., 2012; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b). These NOM constituents, together with other abiotic and biotic constituents, accumulate along the vertical depth of the RGS filter bed with increased operating time since the RGS filter was backwashed, which consequently increases the head loss of the RGS filter, which may lead to clogging and hampers the RGS filter's performance (Liu *et al*., 2012).

It was speculated above that three unknown gammaproteobacterial species of the genera *Aeromonas* (OTU5) and *Acinetobacter* (OTU12 and OTU18) were indigenous inhabitants of the RS filter that can attach, persist and proliferate in the void spaces of the RS filter particles, as their relative abundances were between 4-fold and 23-fold higher with operating time since the RS filter was backwashed (that is, 34 hours after the RS filter has been backwashed). In relation to this study, these unknown gammaproteobacterial OTUs exhibited no base-pair differences in their V4 16S rRNA sequences, when compared to the three dominant unknown gammaproteobacterial species of the genera *Aeromonas* (OTU1) and *Acinetobacter* (OTU2 and OTU3) identified before. Given this, we restate that these *Gammaproteobacteria* species gain a selective advantage in the RS filter bed with increased operating time since the RS filter was backwashed, most probably due to the accumulation of NOM constituents in the RS filter bed.

In addition to *Proteobacteria*, we reported the dominance of *Bacteroidetes* and *Actinobacteria*. Previously, *Bacteroidetes* have been reported as the most abundant phylum in freshwater systems, particularly in the lake epilimnion, where they comprise a large proportion of the particle-associated microbial communities (Nold and Zwart, 1998; Pernthaler *et al*., 2004; Lemarchand *et al*., 2006). Also, species of *Bacteroidetes*, particularly species of the genus *Flavobacterium*, have been reported to be ubiquitous in DWTPs (Zeng *et al*., 2013). In general, species of *Bacteroidetes* have unique metabolic features, including carbon fixation through anaplerotic pathways and rhodopsin-based photometabolism, which attribute to their survival in freshwater systems (González *et al*., 2008). Similar to some *Gammaproteobacteria*, species of *Bacteroidetes* are also copiotrophs, and are capable to degrading complex polymeric NOM (Kirchman, 2002) and easily biodegradable NOM, including DOC (Eiler and Bertilsson, 2007; Zeder *et al*., 2009).

Within *Actinobacteria*, the ACK-M1 clade accounted for most, which was about 70.55% of the *Actinobacteria* sequences. Previously, the ACK-M1 clade has been reported as the most abundant clade in microbial communities found a variety of freshwater systems, ranging from high-nutrient copiotrophic, moderate-nutrient mesotrophic and low-nutrient oligotrophic freshwater systems (Martinez-Garcia *et al*., 2012; Llirós *et al*., 2014). Also, the ACK-M1 clade has been reported to be an ubiquitous inhabitant of DWTPs, particularly within RGS filtration systems (Zeng *et al*., 2013). Reasonable explanations for their ubiquitous distribution in RGS filter systems is not clear; however, their survival in freshwater systems is assumed to be enhanced by their small size and unique metabolic features, including carbon fixation through anaplerotic pathways and actinorhodopsin-based photometabolism, which, also may attribute to their distribution in RGS filter systems (Newton *et al*., 2011; Martinez-Garcia *et al*., 2012; Garcia *et al*., 2013; Ghai *et al*., 2014).

During the second study (DWTP 2), 251 OTUs were identified with the most abundant OTU being affiliated to an unclassified genus with in the *Actinobacteria* phylum. *Pseudomonas* and *Rhodoferax* were the second and third most abundant OTUs. The last two of the top five most abundant OTUs were classified to *Bacteroidetes*, *Flavobacterium* (genus) and *Cyanobacteria*, *Synechococcus* (genus), respectively. This is in contrast to the first study where A*eromonas* and *Acintobacter* were identified as OTU1 and both OTU2 and OTU3, respectively. From this study, RS filter samples were collected only from one filter house, whereas in the study of DWTP 1, three filter houses with different operational and structural features were sampled. It is believed that these factors could introduce more variability into the microbial community. However, *Flavobacterium* was also identified as one of the relatively high abundant genera. In the second study the majority of the OTUs were assigned to the domain Bacteria, of which *Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria* and *Acidobacteria* were the dominant phyla. Regardless of the different treatment processes, multiple studies found these phyla to be dominant in the RS filter microbial communities (Pinto *et al.*, 2012; Li *et al.*, 2017). Interestingly, Pinto *et al.* (2012), found *Nitrospira* and the phylum OD1 to be two of the seven phyla with the highest abundance, whereas in this study these two phyla ranked ninth and seventeenth, respectively, out of the 19 phyla identified. The *Cyanobacteria* was one of the more dominant phyla. *Gammaproteobacteria*, *Betaproteobacteria* and *Alphaproteobacteria* were the dominant classes identified amongst the *Proteobacteria*, which also corresponds with the findings of other studies (Pinto *et al.*, 2012; Li *et al.*, 2017).

3.4.2 Spatial and temporal variations of the rapid sand filter microbial communities

In the initial study, FB samples were collected from nine RS filters located within the filter galleries of three PSs, located in the same DWTP, to determine the extent of spatial variation in the microbial communities found: (i) along the surface and depth of the RS filter bed, (ii) across parallel RS filters within a filter gallery, and (iii) across different filter galleries. Although these PSs are operationally similar, as each treated the surface water through the successive occurrence of six conventional treatment operations i.e. coagulation, flocculation, sedimentation, carbonation, filtration (i.e. RS filtration) and disinfection (i.e. chlorine disinfection). However, each PS functions independently and differs in certain operational and structural features. Some differences include, source water intake (i.e. closed pipe or open canal system), RS filter material age, RS filter size (i.e. 64 m^2, 169 m^2 and 144 m^2 for PS1, PS3 and PS4, respectively), backwash regime, flow rate, and hydraulic retention time. All of these differences, to some extent, can introduce variability at each PS that might influence the microbial communities of the RS filters harboured within each of the filter galleries. However, given these compounding differences and the complexity of the DWTP layout, it may be difficult to identify the precise factors responsible for changes in the RS filter microbial communities.

Based on the sampling of the first DWTP, no spatial differences in the RS filter microbial communities, along the surface and depth of the RS filter bed, and across parallel RS filters within a filter gallery were found, which were supported by AMOVA and parsimony comparisons. These findings were in congruence with previous studies (Bai *et al*., 2013; Lautenschlager *et al*., 2014; LaPara *et al*., 2015) as well as the results from the second study performed during this project (DWTP 2). It is generally accepted that RGS filter systems, including RS filters and GAC filters, retains a homogenous microbial community, as these filter systems are

regularly backwashed to reduce the head loss caused by the accumulation of abiotic (i.e. NOM) and biotic constituents (Liu *et al*., 2012; Lautenschlager *et al*., 2014). Likewise, analogous microbial communities are expected in RGS filter systems harboured within the same filter gallery, as the physical and chemical properties of the influent water that, in general, are known to influence the microbial communities of RGS filter systems are most likely very similar (Li *et al*., 2010; Pinto *et al*., 2012; Liao *et al*., 2013; Kim *et al*., 2014; Albers *et al*., 2015). However, in contrast to these concluding remarks, two recent studies reported local (i.e. within a RS filter) and intra (i.e. between RS filters in a filter gallery) variation amongst RS filter microbial communities (Gülay and Smets, 2015; Gülay *et al*., 2016). Using novel statistical approach proposed by Gülay and Smets (2015), they reported that local and intra variation amongst RS filter microbial communities stems, in a large part, from low-abundant OTUs. Linking to these findings, we found high local and intra dissimilarities in the memberships of the RS filter microbial communities, that is between 60% and 80% dissimilar memberships, attributed by numerous low-abundant OTUs that were exclusively associated with each spatial grouping or sample site, i.e. along the surface and depth of the RS filter bed, and across parallel RS filters within a filter gallery (data not shown). Given our findings, it is plausible to assume that local and intra variation amongst the RS filter microbial communities may exist.

Furthermore, we found differences in the RS filter microbial communities across filter galleries during the first study (DWTP 1). This was supported by AMOVA and parsimony comparisons. In particular, the microbial community of spatial grouping PS1 was significantly different from the microbial communities of spatial groupings PS3 and PS4, in contrast, were not significantly different from each other. Several previous studies concur that selected chemical and physical properties of the influent water may have an impact on the microbial community of RGS filter systems, including temperature (Pinto *et al*., 2012; Kim *et al*., 2014), and altered concentrations individual chemical constituents (i.e. phosphorus, nitrogen, iron and methane) (Li *et al*., 2010; Liao *et al*., 2013; Albers *et al*., 2015) or DOM constituents (i.e. DOC and DON) (Liu *et al*., 2012; Bai *et al*., 2013; Feng *et al*., 2013; Liao *et al*., 2013; Lautenschlager *et al*., 2014; Liao *et al*., 2015b). In this study, the source water that feeds the DWTP (DWTP 1) was extracted from a large, nutrient-rich copiotrophic or eutrophic surface water reservoir, and transported to the DWTP, either through a close pipe system that directly feeds PS1, or an open canal system that flows into a reservoir located within the DWTP, which directly feeds PS3 and PS4 (Ewerts, 2010). Given these differences in source water intake, we speculate that the chemical and physical properties of the source water, though extracted from the same surface water reservoir, may be altered as it is transported towards the DWTP. It will be interesting, in future work, to compare the physicochemical properties of the source water inlets, and subsequently identify correlations between the RS filter microbial community structures and physicochemical water properties. However, only one filter house was sampled at the second DWTP and no comparison between filter houses could be made for this part of the study.

In addition to spatial variability, temporal variation in the RS filter microbial communities, in DWTP 1, was also observed. As determined from the membership-based and structure-based homogeneity of multivariate dispersion tests, the extent of temporal variation in the RS filter microbial communities were similar amongst the PSs. Similar to previous studies, we reported that the RS filter microbial communities of the month

groupings at each PS were different, which was supported by AMOVA and parsimony comparisons (Kim *et al*., 2014, LaPara *et al*., 2015). Several reasonable explanations for these differences have been reported in literature, including changes in the chemical and physical properties of the influent water that are attributed by seasonal changes or process changes in the DWTP on a diurnal, monthly or seasonal basis (Liu *et al*., 2012; Pinto *et al*., 2012; Liao *et al*., 2013; Kim *et al*., 2014; Albers *et al*., 2015; Liao *et al*., 2015b; Gülay *et al*., 2016).

A temporal trend was also observed for both alpha and beta diversity metrics during the analysis of the data collected at the second DWTP. Samples collected during different seasons, showed a temporal variation in the alpha diversity, especially in the richness and diversity, and community structure of the microbial community. This may be linked to the adaptation of the microbial community to the change in source water temperature. In addition, the dissimilarity in community structure increased between all the RS filter samples as the monthly interval increased.

3.4.3 Spatial and temporal trends of the rapid sand filter microbial communities

Spatial and temporal trends of selected microbial OTUs in the RS filters of the filter galleries were visually portrayed using 0.77% of the detected OTUs, which was reported to sufficiently explain the changes in the microbial community structures based on Mantel's test. This small subset of OTUs used to explain the changes was not unusual, as similar explanatory powers of small subset OTUs has been documented in other related and unrelated studies (Fuhrman *et al*., 2006; Pinto *et al*., 2014).

The spatial and temporal trends provided considerable insight into the dynamics of the RS filter microbial communities. At the first DWTP, four OTUs that were detected were classified as unknown species of the genera *Aeromonas* (subphylum: *Gammaproteobacteria*), *Acinetobacter* (subphylum: *Gammaproteobacteria*) and *Flavobacterium* (phylum: *Bacteroidetes*). These bacteria were persistently detected at high relative abundances and collectively constituted, depending on the PS, for between 70% and 80% of the microbial community. Similarly, two OTUs classified as unknown species of the family *Enterobacteriaceae* and genus *Rheinheimera* were detected at relative abundances that were between 4-fold and 20-fold higher in the microbial community of PS1, when compared to the microbial community of PS3 and PS4. The presence of these OTUs at higher relative abundances, along with the presence of numerous rare abundant OTUs (which are likely OTUs from previous colonization events on the RS filter bed, or independent OTUs that have detached from biofilms in the surrounding environment) are most likely responsible for the inter (i.e. between filter galleries) variation that were reported amongst the RS filter microbial communities (Aizenberg-Gershtein *et al*., 2012; Pinto *et al*., 2012; Zeng *et al*., 2013; Lin *et al*., 2014);.

Moreover, given the high relative abundances of these OTUs, it is plausible to assume that they are indigenous inhabitants of the RS filter that have the ability to attach, persist and proliferate in the void spaces of the RS filter particles. Also, as stated previously, these OTUs likely represent the core RS filter microbial community that are possibly involved in biological mediated processes, which aids in improving the quality of water (Gülay *et al*., 2016). As discussed earlier, the microbial species previously reported to be involved in

biological mediated processes in RGS filter systems were found at considerable low relative abundances, while the precise biological functions of the dominant RS filter OTUs are currently not known. In future work, it will be interesting to elucidate the exact biological function of the dominant RS filter OTUs.

3.4.4 Local distance-decay relationships

The RS filter microbial communities exhibited local (i.e. within RS filter) succession patterns, which were visually represented in membership- and structure-based distance-decay relationships. At the first DWTP, the distance-decay relationships demonstrated significant positive correlations between differences in the memberships and structures of the RS filter microbial communities, at distances along the surface and depth of the RS filter bed, which intuitively implies that the microbial community of the RS filter changes over spatial scales (Hanson *et al*., 2012; Nemergut *et al*., 2013). Distance-decay relationships of microbial communities have previously been reported in numerous environmental settings, including aquatic (i.e. surface water and groundwater settings and marine settings) and terrestrial settings (Redford and Fierer, 2009; Bell, 2010; Youssef *et al*., 2010; Ayarza and Erijman, 2011; Soininen *et al*., 2011; Zinger *et al*., 2011; Pinto *et al*., 2014; Gülay *et al*., 2016). However, to our knowledge, this is the first report of distance-decay relationships in RS filtration systems.

A similar distance-decay relationship observed between the different depths of the RS filter sampled at the second DWTP. This indicated that the further the two samples are from one another the more dissimilar they are. This is to be expected as sand filtration, or any other biofiltration processes for that matter, is incorporated into the treatment process of drinking water to effectively remove organic and inorganic substrates, turbidity and microorganisms through the biofilms that grow on the granular medium (Pinto *et al.*, 2012; Proctor and Hammes, 2015).

The exact mechanisms that govern distance-decay relationships are not clear. However, in a review by Nemergut *et al*. (2013), they suggest that the conceptual synthesis of community ecology proposed by Vellend (2010), which consist of four ecological processes (i.e. drift, dispersal, speciation, and selection) can be used as a theoretical framework to understand the mechanisms governing distance-decay relationships. However, given the complexity of these processes, it might not be possible to identify the precise mechanisms responsible for the local distance-decay relationships that were observed amongst the RS filter microbial communities.

As reported in recent literature, several factors can influence the microbial communities of RGS filtration systems, including treatment operations prior to RGS filtration (Fonseca *et al*., 2001; Lautenschlager *et al*., 2014), physical and chemical water properties (Li *et al*., 2010, Pinto *et al*., 2012; Liao *et al*., 2013; Kim *et al*., 2014; Albers *et al*., 2015) and periodic backwashing (Kasuga *et al*., 2007; Liu *et al*., 2012; Gibert *et al*., 2013; Liao *et al*., 2015b). Reasonable explanations for the distance-decay relationships across the surface of the RS filter bed are not clear. However as mentioned above, distance-decay relationships along the depth of the RS filter bed likely exist due to the preferential attachment and detachment of selected microbial species that are

either removed along the depth of the RS filter bed or infiltrates along the depth of the RS filter bed, to be subsequently seeded from the RS filter bed into downstream treatment operations (Pinto *et al*., 2012, Lautenschlager *et al*., 2013, Zeng *et al*., 2013, Lautenschlager *et al*., 2014, Lin *et al*., 2014, Liao *et al*., 2015a, Li *et al*., 2017). From our data, we gathered some supporting evidence for these explanations. Firstly, along the depth of the RS filter bed, the microbial diversity persistently declined, with an approximate 1.00-fold, 1.25-fold, 1.30-fold, and 2.00-fold decrease in richness 5 cm, 10 cm, 20 cm and 30 cm from the upper layer of the RS filter bed, respectively. This decreasing observation intuitively implies that some microbial species are removed along the vertical depth of the RS filter bed (Wang *et al*., 1995). However, implementation of a quantification measure is required for a more accurate explanation thereof. Secondly, as reported in our previous study, we identified some RS filter independent microbial species that were seeded from the RS filter bed into the downstream treatment operations.

3.5 SUMMARY

This study provides a detailed interpretation on the membership and structure of microbial communities that inhabits RGS filter systems located within large-scale DWTPs. The RS filters harboured complex microbial communities that were preferentially dominated by copiotrophic OTUs, which likely attach, persist and proliferate in the void spaces of the RS filter particles. We reported spatial (i.e. local, within RS filter; intra, within filter gallery; and inter, between filter galleries) and temporal variation in the membership of the RS filter microbial communities that were mainly attributed to low-abundant OTUs. In addition we reported on variation in the structure of the RS filter microbial communities that were, also, attributed to variability in the relative abundances of the dominant OTUs.

Although some level of local and intra variation amongst the RS filter microbial communities might exist, the beta diversity data based on comparisons within and between filters indicated that no significant spatial differences in the RS filter microbial communities (i.e. along the surface and depth of the RS filter bed, and across parallel RS filters within a filter gallery) were present. This data indicates that the sand filter community is homogenously distributed across the filter at a specific point in time as well as between filters within the same filter house. The implications of this finding is that the dominant members of the filter community within a filter house could be based on a representative sample taken from one of the filters in the filter house. This has definite implications for efforts to predict the microbial community of the distribution system based on microbial community associated with the filter bed.

CHAPTER 4: BACTERIAL COMMUNITY DYNAMICS WITHIN A BULK WATER DISTRIBUTION SYSTEM

4.1 INTRODUCTION

Despite all efforts taken to remove microorganisms during the purification process, the final microbial concentration of treated drinking water can still be between 10^6-10^8 cells per litre after disinfection (Kahlisch *et al*., 2012; Pinto *et al*., 2012; Chiao *et al*., 2014; Holinger *et al.*, 2014; Bautista-de los Santos *et al*., 2016). Further challenges faced by all water utilities are to ensure the biostability of the drinking water once it enters the distribution system (Prest *et al.*, 2016). Maintaining the biostability is often difficult, as several factors impact the composition and growth of the microbial community inherent to the system. Factors influencing the biostability of the drinking water in DWDSs include pipe materials (in both DWDS and plumbing), disinfectant regime, nutrient availability and environmental factors such as temperature and pH (Lautenschlager *et al*., 2013; Shade *et al*., 2013; Pinto *et al*., 2014; Prest *et al*., 2016). The difficulty in controlling the influence of these factors on the microbial community is that all factors, including the microbial community, are interconnected and may result in different impacts depending on the combination in which they occur.

In most drinking water distribution systems, the disinfection regime, i.e. the type of disinfection agent, the concentration used, point of administration and potential depletion thereof in the system, is the main factor impacting on the microbial ecology of the system (Mains, 2008; Liu *et al*., 2013; Wang *et al*., 2014; Prest *et al*., 2016). Chlorine and chloramine are the two most commonly used disinfecting agents. Chloramine is better at penetrating biofilms, whereas chlorine is less effective as it is consumed by biofilms through side reactions with organic material. Therefore, chloramine is now widely used as a residual disinfectant as it stays active much longer than chlorine (Mains, 2008).

Environmental factors such as temperature and pH were also shown to have an important impact on the microbial ecology of drinking water. Temperature alters microbial growth kinetics and can amplify the competitive interaction between microorganisms in the community with some species out-competing others at higher or even lower temperatures. Seasonal patterns in the microbial community have primarily been linked to temperature changes in the source water (Liu *et al*., 2014; Pinto *et al.*, 2014; Prest *et al*., 2016). Bacterial concentrations have been shown to be highest in the summer seasons, which can potentially cause health and aesthetic problems for consumers (Prest *et al*., 2016). Seasonal temperatures also have an effect on microbial mediated processes such as nitrification as it influences the growth of the nitrifying bacteria (Wang *et al*., 2014).

Apart from the physicochemical factors, the microbial community residing in the drinking water can also impact the biostability of the drinking water, thereby influencing the water quality. The impact varies depending on the different species of microorganisms present in the drinking water, their interaction with one another in the different phases of drinking water (i.e. biofilm, sediments and bulk water) and the biologically processes they carry out (for example nitrification and bio-corrosion). The different water phases, i.e. bulk water, biofilms, loose deposits (sediment) and suspended solids harbour diverse microbial communities as each phase provides a different niche (Henne *et al.*, 2012; Liu *et al.*, 2013; Liu *et al.*, 2014' Prest *et al.*, 2016).

The majority of the microorganisms typically detected in drinking water distribution systems are known to inhabit aquatic environments (Kahlisch *et al.*, 2012). These include bacteria from the phylum *Proteobacteria*, especially *Beta-, Alpha-, and Gammaproteobacteria* (Emtiazi *et al.*, 2004; Henne *et al*., 2012; Kahlisch *et al.*, 201; Lautenschlager *et al*., 2013; Zarraonaindia *et al*., 2013; Mi *et al*., 2015; Bautista-de los Santos *et al*., 2016). Other microorganisms found to dominate the drinking water community are members of *Bacteroidetes, Actinobacteria* (Zarraonaindia *et al*., 2013), *Mycobacteria* (Gomez-Smith *et al*., 2015), *Cynaobacteria, Firmicutes* (Mi *et al*., 2015), *Planctomycetes, Chloroflexi*, *Nitrospira* and a large group of unclassified microorganisms (Emtiazi *et al*., 2004; Henne *et al*., 2012; Kahlisch *et al*., 2012; Pinto *et al*., 2014). Furthermore, from their study, Liu and colleagues (2014) identified bacteria belonging to the genera *Nitrospira, Nitratireductor*, and *Chloroflexus*, specifically inhabiting loose deposits.

Apart from the dominating microorganisms, there are certain undesired microorganisms that are potentially found in drinking water distribution systems. These undesired microorganisms are often involved in microbial mediated processes such as nitrification, iron-oxidation and sulphate-reduction. These processes lead to the undesirable depletion of disinfectant residuals and corrosion of pipe material, respectively. This ultimately causes a decline in the drinking water biological stability and quality. Other undesired microorganisms are hygienically relevant opportunistic pathogens that are associated with health risks (See Chapter 6).

The understanding the microbial ecology of DWDSs has been greatly improved due to the development of culture-independent methods. The use of a set of recently developed culture-independent approaches provides microbial ecologists with a more detailed understanding of the spatial and temporal changes in the microbial communities. These techniques also have great potential to improve our understanding of the diversity and impact of microbial communities associated with drinking water during all stages, from the treatment plant, through the distribution system, up to the consumer's tap.

The aim of this part of the study is to:

- To determine the bacterial community composition at different sampling points within the treatment plant and associated distribution system by means of 16S profiling.
- To compare the community composition at specific sampling points within the treatment plant and water distribution system over time.

- To compare changes in the composition of the communities associated with the treatment plant and the distribution system.
- To determine the level of bacteria within the system by means of flow cytometry.
- To determine the core microbial community within the system and to correlate their occurrence with water quality parameters.

4.2 MATERTIALS AND METHODS

4.2.1 Sampling sites

The DWTP studied during this part of the study was responsible for the daily treatment/purification of 1800 mega-litres (ML) of source water, extracted from a large dam, acting as the main water source for the area. The treatment of the source water follows a conventional treatment system, which includes coagulation, flocculation, sedimentation, stabilisation, rapid sand filtration and disinfection. During coagulation and flocculation, a combination of hydrated lime (55-70 mg/L calcium oxide) and ferrichloride (1-5 mg/L ferrichloride) are used as the primary coagulant along with activated sodium silicate (1-3 mg/L silicon dioxide), which also aids in flocculation. Ferrichloride is used again as a secondary flocculant which further aids in filtration. Due to the addition of hydrated lime, the pH of the treated water increases to a pH of between 10 and 11, which aids in the removal of organic and inorganic (i.e. heavy metal) substances and microorganisms as well as limits the growth of algae. During sedimentation, the flocks settle out in specific horizontal flow sedimentation tanks. The resulting sludge that forms contains various inorganic elements, including complex silicate, aluminium and iron, calcium carbonate and magnesium hydroxide. During stabilisation, the pH of the water flowing from the sedimentation process is decreased from 10.5 (due to the lime coagulant) to 8.0-8.4 to prevent the corrosion and scaling of the pipes within the system. This occurs at the carbonation bays, where a combination of carbon dioxide gas and air is bubbled through the water. After stabilisation, the water is passed through a rapid gravity sand filtration system, where remaining particles in the water were removed.

Disinfection includes the addition of chlorine as a primary disinfectant, followed by the addition of chloramine as a secondary disinfectant. As there are no chlorine contact chambers, disinfection with chlorine occurs within the distribution pipelines along the route to the booster station. The chlorine dosage added to the water, depended on the quality of the raw water, but generally the dosage fluctuates between 1-4.0 mg/L resulting in 1-2.5 mg/L free chlorine residual after a contact time of 20 minutes. Secondary disinfection occurs at a booster station, where ammonia is added to the chlorinated water in a ratio of 4:1 chlorine and ammonia, respectively. This dosing ratio allowed for a concentration of 0.8-1.5 mg/L monochloramine to be maintained throughout the distribution system up to the end-point at the consumer's tap. After chloramination at the booster station, the drinking water is pumped via the distribution system to an urban area about 190 km from treatment plant where it was used to supply multiple households and industries with drinking water.

Collection of both water and sand samples, from the filter bed media, was conducted on a monthly basis for one year (February 2016-January 2017). Specifically, the filter bed media (Sand) samples were collected from the middle of a single designated rapid gravity sand filter (RSF). Furthermore, the water samples were collected from designated sampling points (Figure 3.1) within the treatment works including source water (RW), filter influent (FI), filter effluent (FE), chlorinated water (Cl), at the booster station before chloramination (BfAm) and after chloramination (ClAm) as well as at five additional points along the distribution system (DSP_1 to DSP_5).

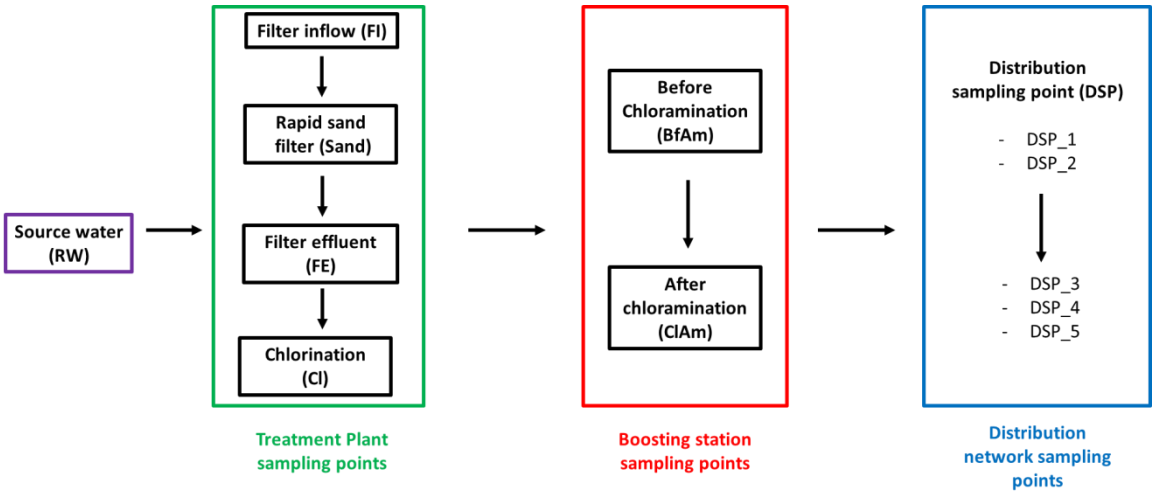


Figure 4.1. A schematic layout of the treatment works and DWDS points sampled during this study.

4.2.2 Sample preparation and processing

Water samples were collected in sterile 1 L and 8 L Large Narrow Mouth Nalgene polycarbonate bottles (Thermo Scientific™, South Africa). A Gilson Minipuls 3 peristaltic pump was used to pump the water samples through a STERIVEX™-GP 0.22 µm polycarbonate membrane filter unit (Merck Millipore, South Africa) for microbial biomass collection. After filtration the polycarbonate filter membranes were stored at -20°C until DNA extraction.

For the sand samples from the filter bed, 8 g of the sand sample was re-suspended in 40 ml distilled water followed by sonication for 30 seconds to facilitate the release of bacterial cells from the sand particles. Following sonication, the liquid suspension was filtered through a STERIVEX™-GP 0.22 µm polycarbonate membrane filter unit (Merck Millipore, South Africa) for microbial biomass collection. After filtration the polycarbonate filter membranes were also stored at -20°C until DNA extraction.

4.2.3 Flow cytometry

Quantification of bacterial cells was determined according to the protocol described by Berney *et al*. (2007). Briefly, the two nucleic acid stains, SYBR Green I and Propidium idodide (PI), were used based on their membrane permeability properties as they give an indication to the total cell counts as well as to the number of intact and damaged bacterial cells. A duplicate set for both water and sand samples were prepared. Each set contained 500 µl of the processed sand and water samples. The first set (stained with 5 µl SYBR Green I) was used to quantify the total cell count, whereas the second set (stained with a combination of 5 µl SYBR Green I and PI) was used to quantify the number of intact cells. The sand samples were diluted in a ratio of 1:10. After staining, the samples were incubated for 13 min in the dark at 37°C. Three additional control samples were prepared. The first control was 500 µl of nuclease free water and the second and third controls were bottled drinking water stained with SG and SGPI, respectively.

4.2.4 DNA extraction

Polycarbonate filter membranes with collected microbial biomass were cut up using an ethanol sterilised scalpel and transferred into a 2 ml Lysing Matrix Tube E (MP Biomedical, South Africa) using an ethanol sterilised tweezer. A modified phenol-chloroform DNA extraction procedure as described in Section 3.2.2 was used. Following extraction, the resulting DNA concentrations were measured using the Nanodrop NA-1000™ Spectrophotometer (Thermo Scientific™, South Africa). To visually inspect the quality of the extracted gDNA agarose gel electrophoresis (1% agarose gel) was performed. The gel was run for 25 min at 90 V and 400 mA.

4.2.5 16S rRNA gene amplification and Illumina MiSeq sequencing

The same procedure as described in Section 3.2.2 was followed. All samples were sequenced at the University of Michigan Medical School (Ann Arbor, United States of America) using the Kozich pair-ended sequencing protocol (Kozich *et al*., 2013).

4.2.6 Sequence processing and data analysis

MOTHUR sequence processing

The sequence and data processing were conducted using MOTHUR according to the procedure described in Section 3.2.3

Alpha and beta diversity analysis

Alpha and beta diversity indexes and the comparison thereof were also performed as described in Section 3.2.3.

Statistical analysis

For statistical analyses MOTHUR and R were used as described in Section 3.2.3.

4.3 RESULTS

4.3.1 Sequencing data

Before reporting on the sequencing data results, 27 of the 137 DWDS samples that were collected in this one year study were removed from the dataset prior to processing, due to failed sequence runs. These DWDS samples included samples collected from the DWTP, i.e. one source water samples (RW_8), three chlorination samples (Cl_3, Cl_8 and Cl_11), five samples after chlorination but before chloramination (BfAm_1, BfAm_6, BfAm_7, BfAm_8, BfAm_12) and four samples after the addition of chloramination (ClAm_1, ClAm_8, ClAm_11, ClAm_12). The remaining 14 DWDS samples were those collected from the distribution system, including three samples from the first distribution sampling point (DSP), two samples from DSP_2, two samples from DSP_3, five samples from DSP_4 and two samples from DSP_5.

Prior to quality filtering thought the MOTHUR pipeline, a total number of 3 044 203 16S V4-region sequences where generated by the Illumina MiSeq platform. Following quality filtering, 2 226 434 sequence reads were retained (Table B1). These sequence reads where identified, classified and grouped into 17 181 Operational Taxonomic Units (OTUs) with a similarity cut-off of 97%. Prior to alpha and beta diversity analyses, the samples were subsampled 1000 times and the minimum number of reads in a sample (n = 512), to ensure that all samples had the same amount of sequence reads when compared to one another during the analyses.

Furthermore, a small subset of OTUs were retained following the removal of OTUs with low sequence abundances based on the total number of sequences of the original dataset, using a cut-off level of 99% (MultiCoLA, Gobet *et al*., 2016) and by the percentage of samples in which they were detected (i.e. detection frequency threshold ≥ 75%). This resulted in the retention of 75 OTUs out of the 17 181 OTUs identified. Interestingly, this small subset of OTUs were identified as a group of dominant microorganisms that were able to represent and indicate the dynamics of the entire DWDS microbial community associated with spatial and temporal changes.

4.3.2 Community diversity

Temporal variation

The alpha diversity metrics were used to determine the richness (observed species, Sobs), diversity (Shannon diversity, H') and evenness (Pielou's evenness, J) of the microbial community within each DWDS sampling point. Along with these alpha diversity metrics, a one-way analysis of variance (ANOVA) test and post-hoc HSD Tukey test were performed, in order to determine whether difference within the microbial communities of the DWDS samples were significant based on a p value < 0.05. Due to subsampling (n = 512), the average Good's coverage across all sampling points was 0.9 ± 0.7, indicating a 90% representation of all OTUs in each DWDS sample.

Based on the alpha diversity metrics (Table 4.1 and 4.2), no significant differences were observed within the richness, diversity and evenness of the microbial community within the DWDS when samples were grouped according to the month in which they were collected (one way ANOVAs, all with p-value > 0.05) (Table 4.3, Table 4.4). Although the overall temporal variation associated with the microbial communities was not significantly different, small variations were noticed on a smaller scale, i.e. between months or seasons. These variations occurred within the average richness between the months whereas the average diversity and evenness of the microbial community stayed relatively constant. Interestingly, the mean richness for the month of February, 88.62 ± 25.48, was lower than the other two summer months, December (Sobs = 100.91 ± 30.49) and January (Sobs = 105.48 ± 43.98) (Table 4.1). A similar pattern was observed when months were grouped according to the respective seasons (Table 4.2). A clear increase in the richness was observed from the colder seasons (autumn: 89.06 ± 34.42, winter: 93.71 ± 40.9) to the warmer seasons (spring: 104.15 ± 57.55, summer: 97.13 ± 32.08). Between the months of the same season (except for spring) variation in the average richness was also evident.

Table 4.1 The average alpha diversity data (richness, diversity and evenness) of the DWDS sampling sites according to monthly groupings (February 2016-January 2017)

Month	Observed Species (Sobs)	Shannon Diversity Index (H')	Pielou's Evenness (J)
February	88.62 ± 25.48	3.2 ± 0.67	0.71 ± 0.11
March	83.46 ± 31.05	3.21 ± 0.59	0.73 ± 0.08
April	95.46 ± 28.26	3.39 ± 0.52	0.75 ± 0.08
May	89.4 ± 44.90	3.07 ± 1.22	0.68 ± 0.21
June	89.97 ± 27.19	3.29 ± 0.81	0.73 ± 0.15
July	96.19 ± 28.44	3.54 ± 0.72	0.78 ± 0.12
August	95.35 ± 62.68	3.21 ± 0.80	0.72 ± 0.11
September	100.94 ± 28.17	3.66 ± 0.38	0.8 ± 0.04
October	91.37 ± 37.99	3.07 ± 0.98	0.68 ± 0.17
November	119.43 ± 86.27	3.58 ± 0.86	0.77 ± 0.09
December	100.91 ± 30.49	3.35 ± 0.57	0.73 ± 0.08
January	105.48 ± 43.98	3.45 ± 0.88	0.74 ± 0.12

Abbreviation: Observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness).

Table 4.2 The average alpha-diversity data (richness, diversity and evenness) of the DWDS sampling points according to seasonal groupings (February 2016-January 2017)

Season	Observed Species (Sobs)	Shannon Diversity Index (H')	Pielou's Evenness (J)
Autumn	89.06 ± 34.42	3.22 ± 0.80	0.72 ± 0.13
Winter	93.71 ± 40.9	3.34 ± 0.77	0.74 ± 0.13
Spring	104.15 ± 57.55	3.42 ± 0.82	0.74 ± 0.12
Summer	97.13 ± 32.08	3.32 ± 0.67	0.73 ± 0.10

Abbreviation: Observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness).

Table 4.3 Analysis of Variance (ANOVA) significance test of the richness, diversity and evenness between the DWDS samples according to monthly groupings (February 2016-January 2017)

Observed Species (Sobs)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	12	12584	1048.7	0.57	0.86
Within groups	97	178870	1844		
Total	109	191454			
Shannon Diversity Index (H')	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	12	4.04	0.34	0.54	0.88
Within groups	97	60.01	0.62		
Total	109	64.05			
Pielou's Evenness (J)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	12	0.12	0.01	0.66	0.79
Within groups	97	1.52	0.02		

Abbreviation: Observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness); degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic (FST).

Table 4.4 Analysis of Variance (ANOVA) significance test of the richness, diversity and evenness between the DWDS samples according to seasonal groupings (February 2016-January 2017)

Observed Species (Sobs)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	3	3352	1117.33	0.63	0.6
Within groups	106	188102	1774.55		
Total	109	191454			
Shannon Diversity Index (H')	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	3	0.55	0.18	0.31	0.82
Within groups	106	63.5	0.6		
Total	109	64.05			
Pielou's Evenness (J)	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Between groups	3	9.57×10^{-3}	3.19×10^{-3}	0.21	0.89
Within groups	106	1.64	0.02		
Total	109	1.65			

Abbreviation: Observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness); degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic (FST).

The beta diversity metrics were used to investigate variations that might occur within the microbial community structure, i.e. focusing on presence, absence and abundance data, as well as on the microbial community membership, which solely focuses on presence and absence data. The metrics used to investigate the community structure were Bray-Curtis pairwise and weighted UniFrac metrics and those used for the community membership were Jaccard and unweighted UniFrac metrics. Along with these four metrics, Analysis of Molecular Variance (AMOVA) was also performed to validate the significance of the variation between the microbial communities within the DWDS samples based on a p-value < 0.05.

Based on seasonal groupings, the majority of the pairwise comparisons for all metrics were not significantly different (AMOVA, p > 0.05). Pearson's correlations also showed a negligible positive correlation (with the exception of the weighted UniFrac metric) between the difference season and community structure as well as community membership (Bray-Curtis Pearson's R = 0.02 & p-value = 0.23; Weighted UniFrac Pearson's R = -0.007 & p-value = 0.58; Jaccard Pearson's R = 0.03 & p-value = 0.01; Unweighted UniFrac Pearson's R = 0.02 & p-value = 0.06). Although the overall seasonal groupings showed no significant differences in the microbial community, a significant difference (AMOVA p < 0.05) was observed between winter and autumn (Bray-Curtis p-value = 0.02 & F_{ST} = 2.25; Jaccard p-value = 0.03 & F_{ST} = 1.53; Weighted UniFrac p-value = 0.04 & $_{FST}$ = 2.03) and between summer and autumn (Bray-Curtis p-value = 0.005 & F_{ST} = 2.92; Jaccard p-value = 0.01 & F_{ST} = 1.67; Weighted UniFrac p-value = 0.02 & F_{ST} = 2.64 and unweighted UniFrac p-value = 0.03 & F_{ST} = 1.61 (Table 4.5).

Table 4.5 Analysis of Molecular Variance (AMOVA) significance test of the richness, diversity and evenness between the DWDS samples according to seasonal groupings (winter compared to summer and autumn)

Bray-Curtis dissimilarity metrics					
Autumn vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.79	0.79	2.25	0.02
Within	61	21.45	0.35		
Total	62	22.24			
Summer vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.97	0.97	2.92	0.005*
Within	51	16.96	0.33		
Total	52	17.94			
Jaccard dissimilarity metric					
Autumn vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.57	0.57	1.53	0.03
Within	61	22.84	0.37		
Total	62	23.42			
Summer vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.62	0.62	1.67	0.01
Within	51	18.83	0.37		
Total	52	19.45			
Weighted UniFrac metric					
Autumn vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.49	0.49	2.03	0.04
Within	61	14.82	0.24		
Total	62	15.31			
Summer vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.59	0.59	2.64	0.02
Within	51	11.48	0.23		
Total	52	12.08			
Unweighted UniFrac metric					
Summer vs Winter	**df**	**SS**	**MS**	**F_{ST}**	**p-value**
Among	1	0.48	0.48	1.61	0.03
Within	51	15.37	0.3		
Total	52	15.85			

4.3.3 Spatial variation

The same alpha diversity metrics (richness/observed species (Sobs), Shannon diversity (H') and Pielou's evenness, (J)) and significance tests were used to determine whether the DWDS microbiome is influenced spatially. A clear decrease in observed species was noticed as the source water flowed through the treatment plant and disinfection processes. However, following disinfection, an increase in richness was observed in the distribution system samples following chloramination (Table 4.6). More specifically, the site locations showing significant different (ANOVA and post hoc Tukey test, $p < 0.05$) in richness can be grouped into three groups. The first group mainly involves RW that is different to samples from FE up until the last DSP sample, i.e. DSP_5. The second group includes FI and sand samples having different richness

than Cl, BfAm, ClAm and DSP_2 and finally the third group includes FE being dissimilar to Cl (Table 4.6). When DWDS samples were grouped according to site location, a clear significant difference was observed in the microbial community's diversity (ANOVA and post hoc Tukey test; Shannon diversity, p-value = 2.39 x 10^{-7} & F_{ST} = 5.95), evenness (Pielou's evenness, p-value = 3.59 x 10^{-4} & F_{ST} = 3.51) and especially in the richness (observed species, p-value = 1.17 x 10^{-9} & F_{ST} = 7.89) of the DWDS samples (Table 4.7).

Table 4.6 The average alpha diversity data (richness, diversity and evenness) of the DWDS sampling points over the one-year study

Location	Observed Species (Sobs)	Shannon Diversity Index (H')	Pielou's Evenness (J)
RW	148.60 ± 65.87	3.98 ± 0.57	0.80 ± 0.06
FI	116.99 ± 49.60	3.75 ± 0.44	0.79 ± 0.04
Sand	136.04 ± 21.53	3.95 ± 0.38	0.81 ± 0.05
FE	95.64 ± 13.52	3.58 ± 0.31	0.78 ± 0.05
Cl	62.25 ± 24.86	2.3 ± 0.70	0.63 ± 0.12
BfAm	67.7 ± 29.81	2.72 ± 1.02	0.64 ± 0.19
ClAm	63.77 ± 19.52	2.64 ± 0.97	0.63 ± 0.20
DSP_1	81.56 ± 17.10	3.28 ± 0.29	0.75 ± 0.75
DSP_2	70.09 ± 23.68	2.86 ± 1.03	0.67 ± 0.20
DSP_3	77.88 ± 15.27	3.05 ± 0.53	0.70 ± 0.10
DSP_4	91.07 ± 9.66	3.42 ± 0.22	0.76 ± 0.03
DSP_5	79.95 ± 20.49	3.08 ± 0.69	0.70 ± 0.12

Abbreviation: Drinking water distribution system (DWDS), source water (RW), filter influent (FI), filter effluent (FE), chlorination (Cl), before chloramination (BfAm), chloramination (ClAm), drinking system point 1 (DSP_1), drinking system point 2 (DSP_2), drinking system point 3(DSP_3), drinking system point 4 (DSP_4), drinking system point 5 (DSP_5); observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness).

Table 4.7 Post-hoc HSD Tukey test the richness, diversity and evenness between the DWDS samples indicating where significant differences occurred

	Observed Species (Sobs)	**Shannon Diversity Index (H')**	**Pielou's Evenness (J)**
RW vs FE	7.96×10^{-3}	-	-
RW vs Cl	2.50×10^{-6}	2.22×10^{-4}	0.04
RW vs BfAm	1.87×10^{-4}	7.10×10^{-3}	-
RW vs ClAm	6.98×10^{-5}	2.99×10^{-3}	-
RW vs DSP1	4.57×10^{-3}	-	-
RW vs DSP2	1.34×10^{-5}	4.78×10^{-3}	-
RW vs DSP3	2.41×10^{-4}	-	-
RW vs DSP4	0.02	-	-
RW vs DSP5	2.48×10^{-4}	-	-
FI vs Cl	0.01	7.10×10^{-3}	-
FI vs BfAm	-	-	-
FI vs ClAm	-	0.03	-
FI vs DSP2	0.04	-	-
Sand vs BfAm	2.76×10^{-3}	7.82×10^{-3}	-
Sand vs Cl	6.82×10^{-5}	2.26×10^{-4}	0.02
Sand vs ClAm	1.13×10^{-3}	3.26×10^{-3}	-
Sand vs DSP2	3.63×10^{-4}	5.11×10^{-3}	-
Sand vs DSP3	4.67×10^{-3}	-	-
Sand vs DSP5	5.15×10^{-3}	-	-
FE vs Cl	-	0.03	-

Abbreviations: source water (RW), filter influent (FI), filter effluent (FE), chlorination (Cl), before chloramination (BfAm), chloramination (ClAm), drinking system point 1 (DSP_1), drinking system point 2 (DSP_2), drinking system point 3(DSP_3), drinking system point 4 (DSP_4), drinking system point 5 (DSP_5); observed species (Sobs, richness), Shannon diversity (H', diversity) and Pielou's evenness (J, evenness).

An overall change in the diversity was observed as the water exits the treatment plant and is disinfected. Similar to the changes in richness, after chloramination, an increase in diversity was observed. This was also observed for the evenness of the microbial community. More specifically, those samples, which had a significant difference in the diversity (Shannon diversity, H') of the microbial community involved samples collected at the disinfection points (Cl, BfAm and ClAm) and DSP_2. Using ANOVA post-hoc HSD Tukey test, the diversity of RW compared to disinfection points (Cl, BfAm and ClAm) and distribution point DSP_2, were significantly different (p-values < 0.05) (Table 4.7). Furthermore, the microbial diversity in the FI was significantly different to Cl and ClAm (ANOVA post-hoc HSD Tukey test, p-value < 0.05). Finally, sand samples were significantly different to disinfection points (Cl, BfAm and ClAm) as well as DSP_2 (ANOVA post-hoc HSD Tukey test, p-value < 0.05). The microbial community evenness (J) was only significantly

different between the chlorination samples (Cl) and that of source water (RW) and sand samples with significant p-values of 0.04 and 0.02, respectively.

The same beta diversity metrics (structure-based: Bray-Curtis and weighted UniFrac and membership-based: Jaccard and unweighted UniFrac) and significant tests were used to investigate whether spatial groupings shaped the DWDS microbial community in its structure and membership. As with the alpha diversity findings, there were significant differences (AMOVA $p < 0.05$, for all metrics) within the community structure and membership between the different sampling locations. These results were also supported by distance decay relationships, i.e. with an increase in distance between sample locations, the microbial community becomes more dissimilar (Figure 4.2).

Although the majority of the spatial comparisons were significantly different, there were some sampling points that were not significantly different. Both the microbial community structure (Bray-Curtis, p-value = 0.08 and F_{ST} = 1.56 and weighted UniFrac, p-value = 0.05 and FST =2.08) and community membership (Jaccard, p-value = 0.14 and F_{ST} = 1.14 and unweighted UniFrac, p-value = 0.55 and FST =0.93) of Cl samples and BfAm samples were not significantly different. The same result was observed when BfAm and ClAm were compared (Bray-Curtis, p-value 0.35 and F_{ST} =1.07; weighted UniFrac, p-value = 0.24 and F_{ST} = 1.20 and unweighted UniFrac, p-value = 0.15 and F_{ST} = 1.14). No significant difference was seen in the community structure between ClAm samples and DSP_2 (Bray-Curtis, p-value = 0.09 and F_{ST} = 1.46 and weighted UniFrac, p-value = 0.43 and F_{ST} =0.99) as well as between FI and FE (Bray-Curtis, p-value = 0.19 and F_{ST} = 1.21 and weighted UniFrac, p-value = 0.30 and F_{ST} =1.14). The weighted UniFrac p-value for the comparisons between DSP_5 and DSP_3 as well as DSP_5 and DSP_4 showed that these sampling points also had no significant difference in their microbial community structure based on phylogeny, p-value = 0.19 (F_{ST} = 134) and 0.07 (F_{ST} = 1.61), respectively (Table B2).

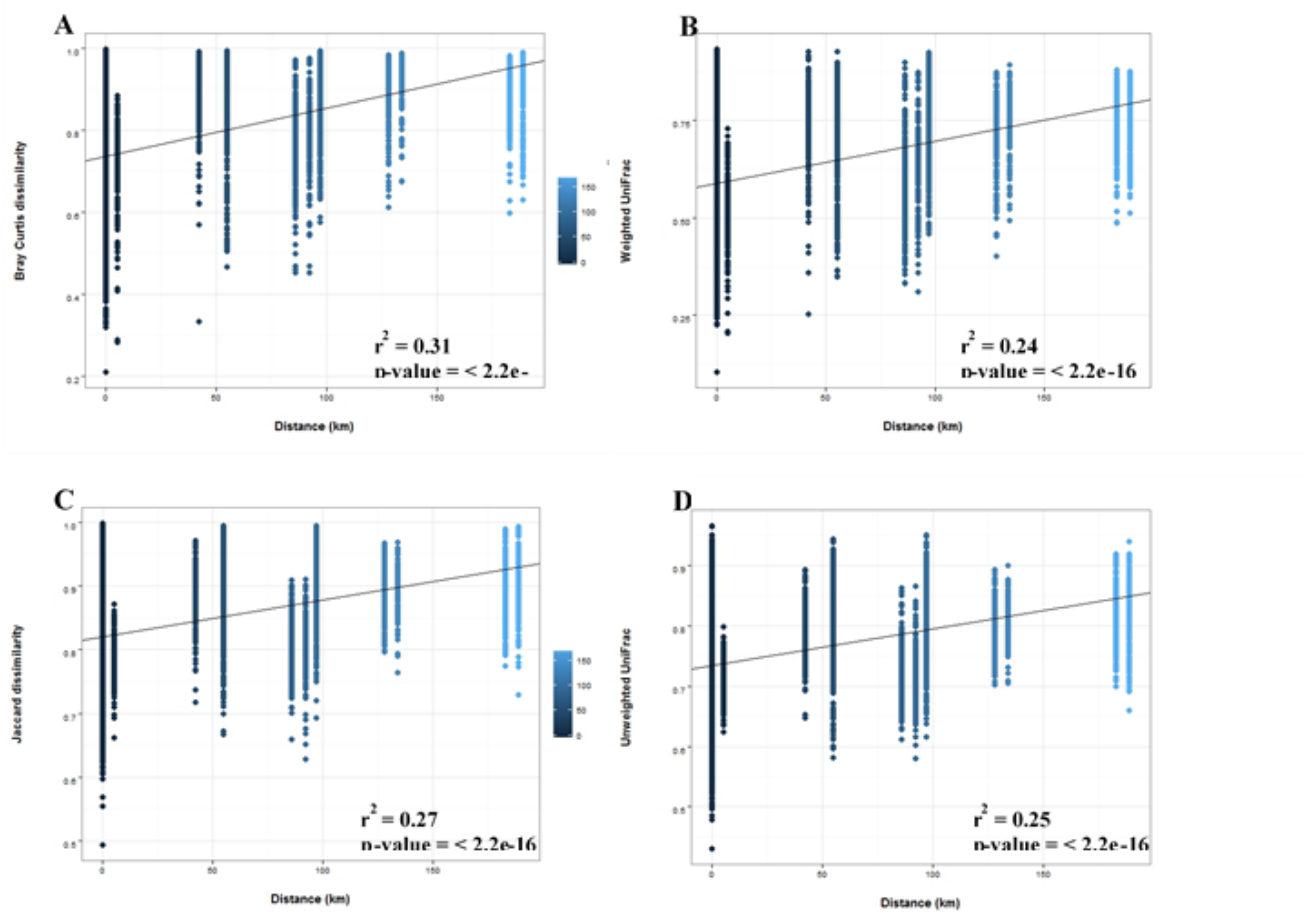


Figure 4.2 Distance-decay relationship showing that as the distance between DWDS sampling point increase is there an increase in dissimilarity in the microbial community structure, (A) Bray-Curtis pairwise distance metric and (B) Weighted UniFrac metrics, and the microbial community membership (C) Jaccard pairwise distance metric and (D) Unweighted UniFrac

4.3.4 Community membership

The top five OTUs were affiliated with only 3 of the dominant phyla including *Actinobacteria* (OTU1), *Proteobacteria* (OTU2, OTU4 and OTU5) and *Planctomycetes* (OTU3). OTU1 had a mean relative abundance (MRA) of 6.47 ± 7.13%, i.e. the average of the relative abundance of that specific classification within each DWDS sampling point. OTU2 was assigned *Betaproteobacteria* class, unclassified genus with a MRA of 3.55 ± 5.21% and OTU3 was classified as belonging to the *Planctomycetia* class and an unclassified genus (MRA = 5.79 ± 10.28%). OTU4 and OTU5 were both identified up to genus level being classified as *Gammaproteobacteria* class, *Pseudomonas* genus with MRAs of 3.50 ± 9.94% and 2.39 ± 9.01%, respectively.

Following the removal of 99% of low-abundance OTUs, the resulting subset of 75 OTUs showed clear spatial patterns that shape the microbial community (Figure 4.3). Considering all the 75 OTUs in the subset dataset, bacteria dominated the microbial community, MRA of 99.05 ± 1.29%, with archaea only having a MRA of 0.95 ± 1.29%. Within the bacteria, 60 phyla were identified of which *Proteobacteria* predominated with a MRA of 49.37 ± 23.34% across all the DWDS samples. *Proteobacteria* was followed by *Actinobacteria, Planctomycetes, Bacteroidetes* and *Acidobacteria* with MRA of 13.87 ± 13.44%, 11.11 ± 14.08%, 9.21 ± 9.50% and 4.72 ± 4.78%, respectively. Within these top five most abundant phyla, 54 classes were present. Of the proteobacterial classes, *Betaproteobacteria* (MRA of 16.920 ± 12.61%), *Alphaproteobacteria* (MRA of 16.13 ± 14.54%) and *Gammaproteobacteria* (MRA of 13.78 ± 17.69%) constituted the top three most abundant classes. The fourth most abundant class was *Actinobacteria* (MRA of 11.25 ± 10.88%), followed by *Planctomycetia* of the phylum *Planctomycetes* (MRA of 9.99 ± 13.69%). Other dominant classes included *Flavobacteria* from the phylum *Bacteroidetes* (MRA of 3.16 ± 6.05%) and *Holophagae* from the phylum *Acidobacteria* (MRA of 2.41 ± 2.89%) (Figure 4.4)

Spatially, *Actinobacteria, Flavobacteria* and *Holophagae* had their highest MRA within the DWTP up until disinfection. Upon disinfection all three of these classes dramatically decreased in MRA and were present in relatively low MRA in the distribution system. *Planctomycetia* however, increased in MRA and remained relatively consistent following disinfection. In addition, after chloramination the proteobacterial classes became more abundant in the DWDS, where both *Beta*- and *Gammaproteobacteria* had a definite increase whereas *Alphaproteobacteria* only showed an increase at the DSP_3-5 sampling site (Figure 4.5).

Although temporal groupings did not have significant effect on the microbial community, all the classes, except those of the proteobacterial phyla and the *Holophagae* class (phylum, *Acidobacteria*), showed changes in their MRA moving from one season to the next. *Actinobacteria* (phylum, *Actinobacteria*) had an increase in MRA in the warmer seasons, whereas *Planctomycetia* (phylum, *Planctomycetes*) had a dramatic increase within the winter months. *Flavobacteria* also had a significant increase in the winter followed by a slight decrease in the spring, which was higher the MRA determined during the summer and autumn (Figure 4.6).

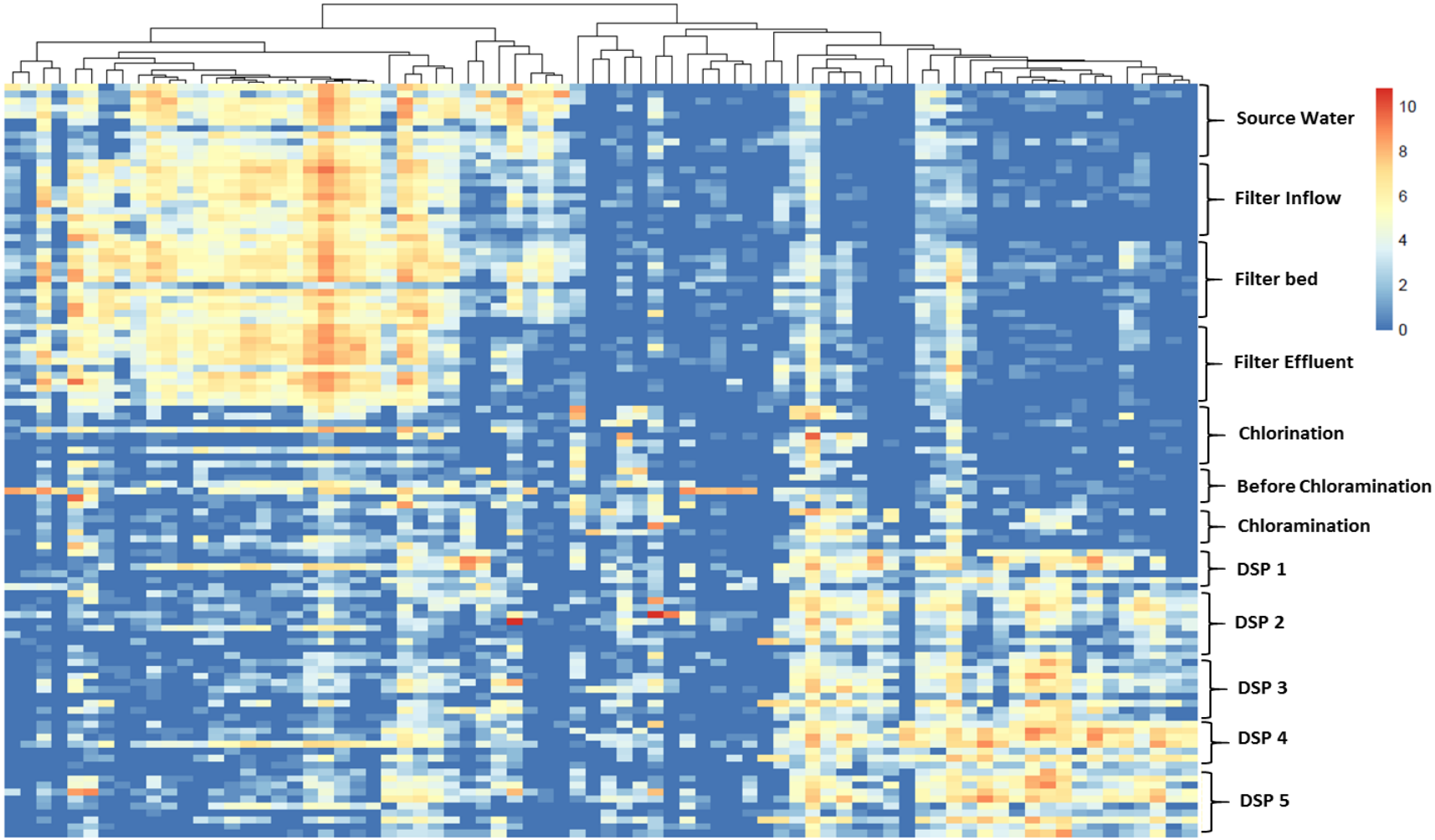


Figure 4.3 Heatmap constructed using the absolute sequence abundances (log transformed) of 75 OTUs that were selected based on their relative abundance within a sample (i.e. relative abundance threshold ≥ 1%) and by the percentage of samples in which they were detected (i.e. detection frequency threshold ≥ 75%). The heatmap boxes were colored from red-to-blue to represent higher-to-lower abundances, and OTUs not represented by a sequence were assigned 10-6 (displayed as dark blue).

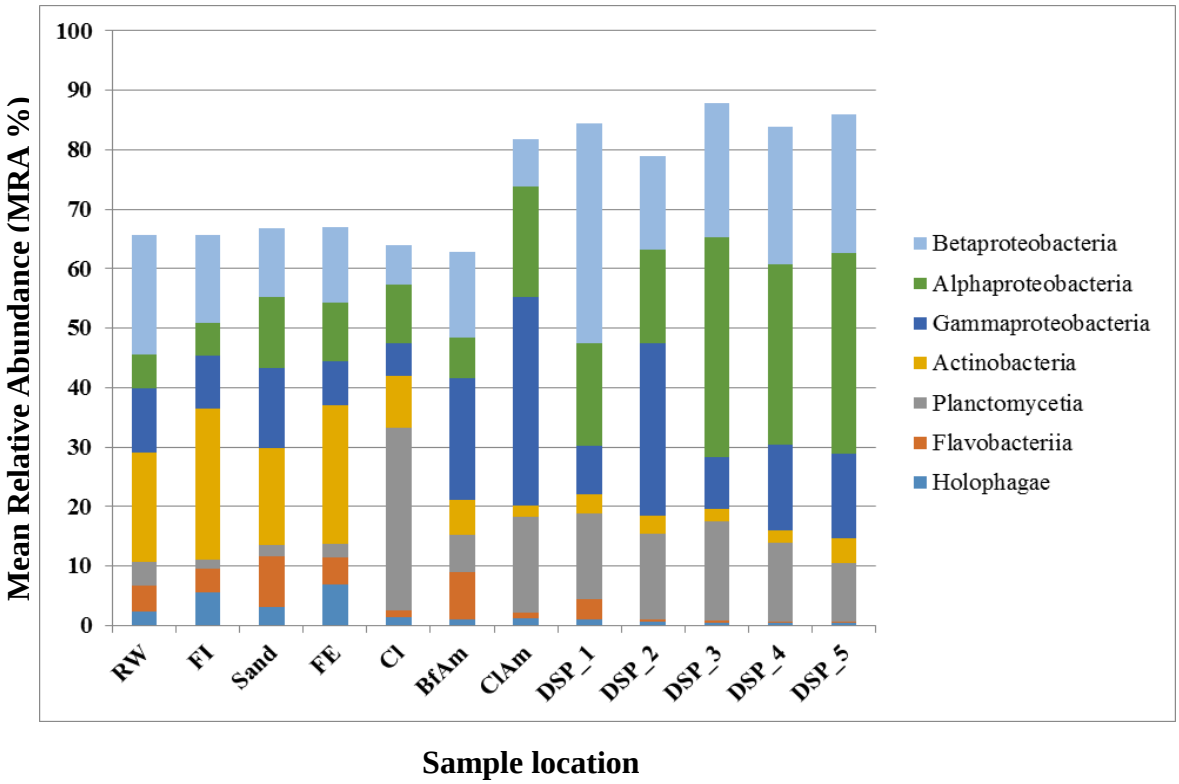


Figure 4.4 Mean relative abundance of the seven most abundant bacterial classes present at all sampling sites within the treatment plant (Source water (RW), filter influent (FI), Rapid Sand Filter media (Sand) and filter effluent (FE)) as well as the distribution system ((Chlorination (Cl), before chloramination (BfAm), chloramination (ClAm), distribution system points (DSP)) over a 12 month period.

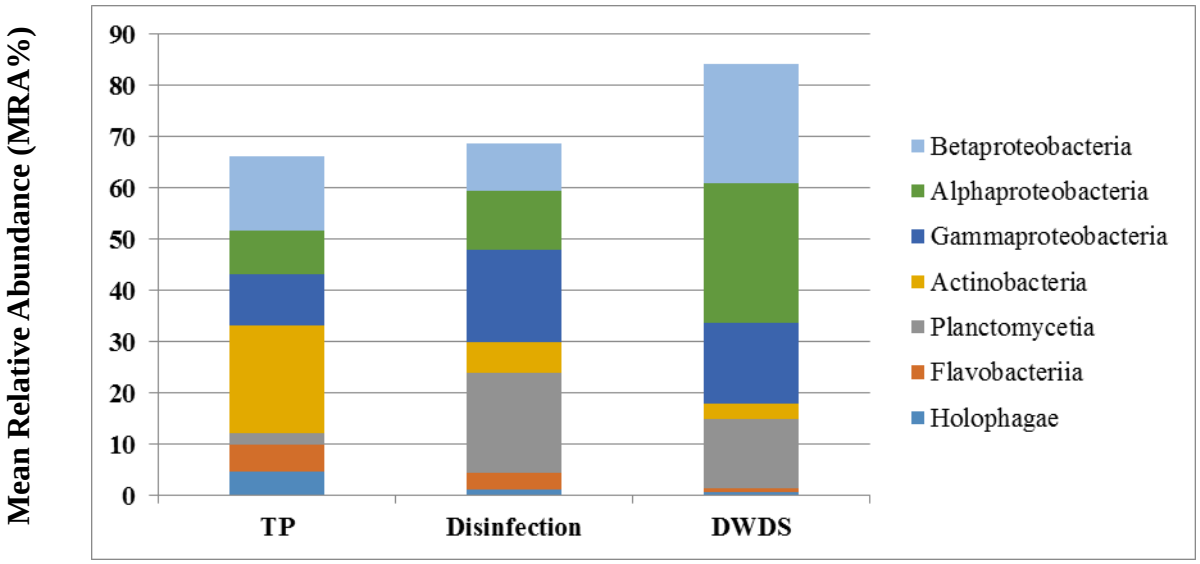


Figure 4.5 Mean relative abundance of the five most abundant bacterial classes in the different stages of treatment (Treatment plant (TP), Disinfection stage, Distribution system (DWDS) over a 12-month period.

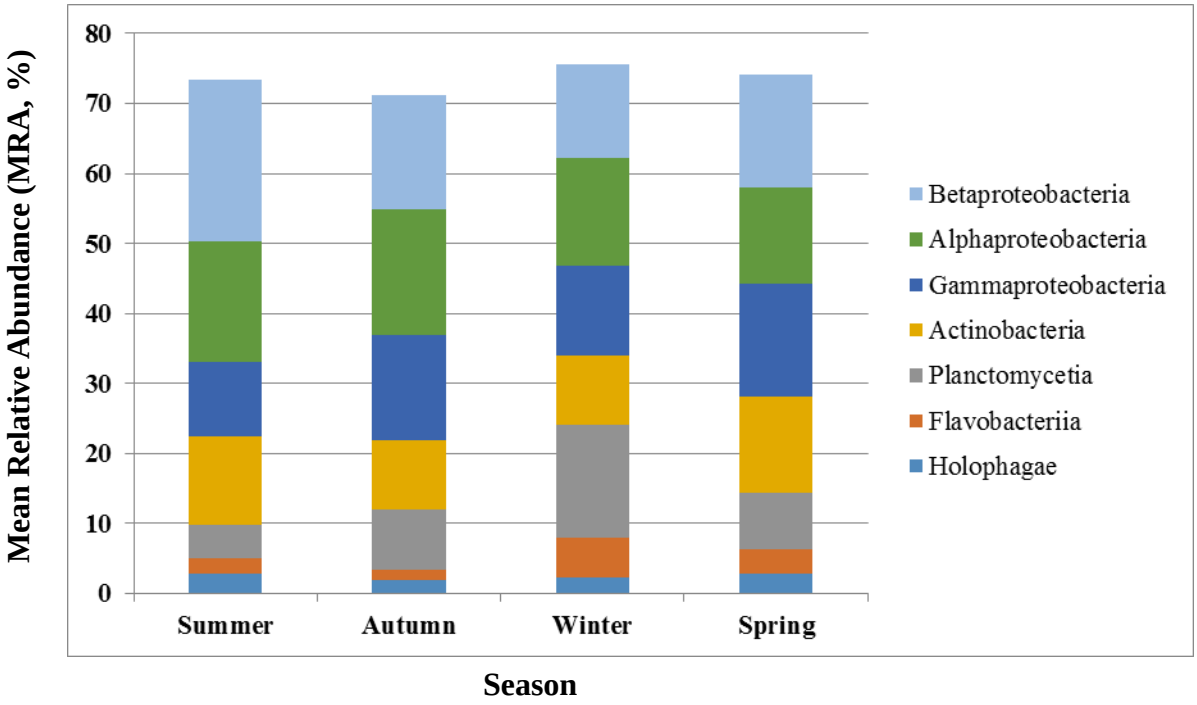


Figure 4.6 Mean relative abundance of the five dominant bacterial classes in DWDS samples showing slight changes due to the temporal groupings.

The 17 181 OTUs identified constituted 4 175 bacterial genera of which only 271 were classified with the remaining 3 904 being unclassified. Unclassified genera were the most abundant in *Actinobacteria* class (MRA = 10.30 ± 11.11%) followed by *Betaproteobacteria*, *Alphaproteobacteria*, *Planctomycetia* and Gammaproteobacteria, which had MRA of 10.09 ± 7.59%, 9.12 ± 11.61%, 8.12 ± 12.08% and 2.55 ± 4.44%, respectively. Of the classified bacterial genera, *Pseudomonas* (*Gammaproteobacteria*), *Nitrosomonas* (*Betaproteobacteria*) and *Sphingomonas* (*Alphaproteobacteria*) were the three dominant genera with MRA of 6.30 ± 13.23%, 2.86 ± 8.52% and 2.75 ± 4.34%, respectively. Within the other phyla, the genera *Flavobacterium* (phylum *Bacteroidetes*, MRA = 2.28 ± 5.51%) and *Planctomyces* (phylum *Planctomycetes*, MRA = 1.713 ± 2.69%) were the dominant genera of within their phyla. Interestingly, the *Nitrospirae* phylum (Class = *Nitrospira*, Order = *Nitrospirales*) had a MRA of 0.57 ± 0.68%, which was not that abundant in comparison to the phyla mentioned and the class *Nitrospira* only had a MRA of 0.38 ± 0.59%.

4.3.5 Flow cytometry

Flow cytometry was performed in order to determine the concentration of both intact and damaged bacterial cells. Based on the average total cell count (TCC) and intact cell count (ICC) across the DWDS sampling points, fluctuations within the cell counts were evident. Decreases in both TCC and ICC were observed as source water moved through the various treatment processes (flocculation, coagulation and sedimentation) and sand filtration (RW, TCC = 2668.92 ± 2833.89, ICC = 1812.24 ± 1381.06; FI, TCC = 1050.64 ± 2115.14 and ICC = 743.78 ± 1429.17 and FE, TCC = 1001.19 ± 1391.84 and ICC = 681.88 ± 923.23). A more dramatic decrease in both counts was observed following disinfection in the chlorinated water (TCC = 175.64 ±296.96

and ICC = 28.11 ± 32.22). However, following chloramination, the TCC and ICC increased by an average of 92.29% and 95.13%, respectively. After this sampling point, yet another decrease in both counts occurs in the DWDS only to increase slightly in the last DWDS sampling point (Figure 4.7).

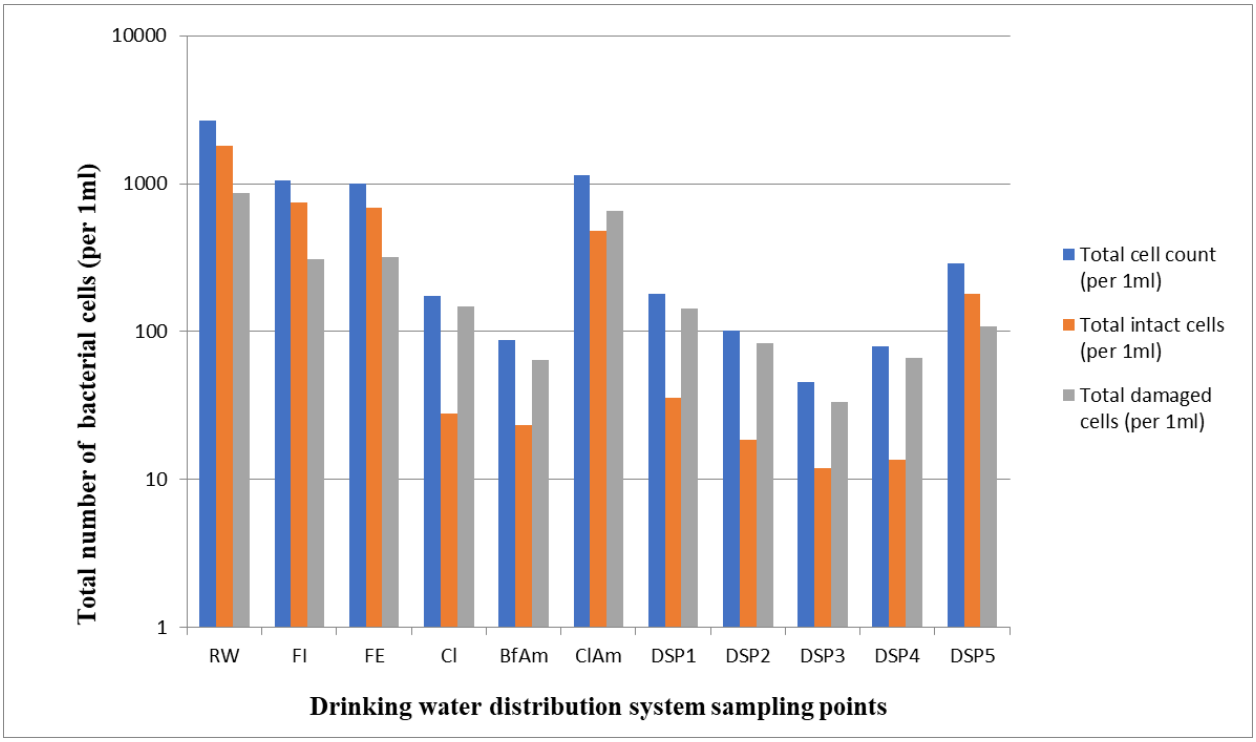


Figure 4.7 The average number of total microbial cell count and intact microbial cell count across all DWDS sampling points.

4.4 DISCUSSION

Overall, no significant differences were observed in the alpha diversity of the DWDS microbial community associated when samples were grouped temporally/seasonally. The one-way ANOVA for all alpha diversity metrics had a p-value > 0.05. Interestingly, looking at the individual sampling sites, temporal variations within the richness of the microbial community occurred. The warmer seasons showed a higher richness compared to the colder seasons. Pinto *et al*. (2014) reported a temporal trend in the richness of the microbial community, with the highest richness recorded in the summer and autumn months. However, they found a strong positive correlation between the microbial community richness and water temperature, whereas in this study the correlation with temporal groupings was very weak. This could be the due to the extreme seasonal changes in the weather typically experienced in Ann Arbor, Michigan, USA where the ratio between surface and groundwater also changes seasonally, with a larger ratio of groundwater used during winter. In contrast these extreme seasonal changes are not typically experienced in South Africa and the source water used is a constant supply of surface water

The beta diversity results further support the weak influence temporal changes had on the microbial community. In addition, Pearson's correlation of all the beta diversity metrics showed a positive yet weak

correlation between an increase in the difference in months between samples and the increase in the dissimilarity of the microbial community. Even though the correlations were positive, the correlation values were closer to 0 than to 1, indicating that temporal groupings are not the main drivers for the variation within the DWDS microbial community.

However, the microbial community of the samples collected in the winter season varied in their community structure and community membership when compared to the microbial community of the DWDS samples collected during summer and autumn. This relatively small influence that temporal changes have on the microbial community, especially in terms of richness, can be attributed to the change in the microbial community of the source water. It is possible that the microbial species diversity within the source water might increase or decrease as the effect changes in seasonal rainfall e.g. with lower rainfall in winter months, the effect of dilution may be limited and the low flow of the rivers during this season may result in an increase in observed species and diversity. The change in richness observed might also be linked to an increase of rare community members in the population.

In contrast, the findings in both the alpha and beta diversity metrics indicated that spatial changes were a stronger driver in shaping the microbial community of the DWDS. Despite the difference in treatment processing, other studies support these findings, as they also concluded that spatial groupings are responsible for the variation observed in the microbial community compared to temporal factors (Proctor and Hammes, 2015; Li *et al*., 2017; Vosloo, 2017). The variations in the microbial community were observed in both the richness and diversity of the community, especially between the DWTP samples and DWDS samples. It was clear that the disinfecting regime brought about the strongest change in the community, as all comparisons having significant differences involved samples collected from the disinfecting sampling sites. Proctor and colleagues (2015) stated that disinfection, which is generally the last purification step, plays a key role in eliminating the vast majority of the microorganisms that persist after filtration, ensuring that potential microbial risk are kept low. The advantage of using chlorine and chloramine as disinfection agents is that these chemical agents may have a long-term effect on the microbial community due to the presence of disinfectant residuals. These disinfectant residuals hinder the growth of certain microorganisms in the distribution system by creating a selective pressure. Each disinfection agent has a different effect on the microbial community, resulting in a change in the microbial composition when added to the water (Proctor and Hammes, 2015). These two disinfecting agents differ in their efficacy in inactivating microorganisms, which is mainly based on their effect on biofilm penetration and the period during which these chemicals stay active. Chlorine is limited to inactivating those microorganisms at the surface of the biofilm whereas chloramine has the ability to penetrate the extracellular polymeric substance matrix (ESP-matrix) of the biofilm (LeChevallier *et al*., 1988; Wang *et al.,* 2014). Chloramine also remains active for a much longer periods than chlorine (Mains, 2008) and reduces the concentration of disinfection by-products.

Additional factors reported to influence the community composition in a disinfected system, is the selection for more resistant microorganisms (Berry *et al*., 2006; Proctor and Hammes, 2015), which potentially leads to unwanted microbial growth in the DWDS. The resistance to the disinfection agents was noticed in the data as

an increase in the microbial diversity after the addition of the disinfection agents was observed, specifically after chloramination. Mi *et al*. (2015) observed this when they conducted a study in which different dosages of chlorine and chloramine were introduced to untreated water. Their focus was on the change in the Proteobacterial community. They found that both *Gamma*- and *Betaproteobacteria* were dominant in those samples which received low and high dosages of chlorine, respectively where as in the chloraminated treated water, *Alpha*- and *Betaproteobacteria* dominated the microbial community. This may be one of the reasons why these three proteobacterial classes were the dominant classes observed within the distribution system after disinfection in this study.

Interestingly, the only significant changes in community evenness were observed for the chlorinated samples when compared to the raw water and filter effluents samples. This again demonstrated the effectiveness of disinfection on the levels of most of the bacterial population and the impact this step has on the DWDS microbial community (Hull *et al*., 2017). There were no significant changes observed in community evenness after the introduction of a disinfectant.

Various other factors may play a role in introducing spatial variation in the microbial community once the water has entered the distribution system, which consists of a number of interconnected pipelines and reservoirs. These factors could influence either the growth or the microbe-microbe interactions that occur within the DWDS. Some of these factors include changes in nutrient concentrations, operational factors, DWDS layout and interconnections, pipeline characteristics as well as several environmental conditions (Berry *et al.*, 2006; Pinto *et al*., 2014; Mi *et al*., 2015; Prest *et al*., 2016). The different phases in the drinking water (i.e. bulk water, biofilm, sediment and loose deposits) differ in their concentration of microorganisms and provide opportunities for various microbe-microbe interactions forming specific niches within the DWDS. The microbes associated with these different niches all have an impact of the microbial community present in the water delivered to the end-user (Proctor and Hammes, 2015). It has been observed that the further the sampling sites are from each other, the greater the influence of the mentioned factors could be, resulting in a higher dissimilarity between the corresponding microbial communities. In this study, a positive correlation between the increase in distance and dissimilarity of the microbiome in the DWDS (distance decay relationship) were observed. However, to pin point which specific factor or parameter has the greater impact on the microbial community remains a challenge as all these parameters work in combination and not independently (Pinto *et al*., 2014).

The microbial community associated with specific sampling sites were similar in structure. Amongst the best examples were the Cl and BfAm samples. After the addition of chlorine (Cl) to the filter effluent water, the treated water exits the treatment plant and flows to the booster station. This similarity in microbial community is highly likely as the water conditions at both these sampling points do not differ significantly, as no major changes in water quality would have occurred between the two sampling points apart from a potential decrease in chlorine residuals in the water. A similar trend was observed between the BfAm sampling site and the ClAm sampling site. Although the water chemistry has changed significantly with the addition of

chloramine, the community may not have had enough time to respond to these changes as the sampling was done directly after the point of chloramination.

Another important observation is the similarity in microbial community structure between the filter influent (FI) and filter effluent (FE) samples and the significant dissimilarity between RSF samples and FI, FE and Cl. This was in clear contrast to the study of Pinto *et al*. (2012) where it was reported that the community of the RSF samples corresponded with samples taken within the distribution system. Their results indicated that microorganisms present on the filter bed were seeded into the DWDS and ultimately shaped the microbiome of the DWDS. In the current study, the filter beds had limited seeding effect and most of the bacteria could be designated as pass-through OTUs, which are defined as OTUs that are not found in the microbiome of the RSF samples, but in both of the filter influent and effluent.

A large number of bacterial species (17 181 OTUs) were associated with the water intended for drinking and household purposes. The three most dominant OTUs belonged to the *Actinobacteria*, *Proteobacteria* and *Planctomycetes* but could not be identified to genus level. It is not uncommon to have unclassified genera present at high relative abundance as the drinking water microbial community is known to harbour a vast number of unclassified microorganisms (Proctor and Hammes, 2015). The function of these bacteria can only be studied using detailed metagenomic studies.

Looking at the overall microbial community, it is clear that bacteria were by far the most dominant and were represented by members of the phyla (in descending order) *Proteobacteria, Actinobacteria, Planctomycetes, Bacteroidetes* and *Acidobacteria*. A number of studies reported the dominance of *Proteobacteria* in the DWDS, with *Beta-, Alpha-* and *Gammaproteobacteri*a as being the most dominant classes (Berry *et al.,* 2006; Pinto *et al*., 2014; Proctor and Hammes, 2015; Hull *et al*., 2017; Li *et al.,* 2017). The high abundance of both *Actinobacteria* and *Bacteroidetes* was also previously been reported (Hull *et al*., 2017; Li *et al*., 2017). Similar to what was found in the study of Pinto *et al*. (2012), *Alphaproteobacteria* increased in MRA from the source water (5.69%) to the last sampling point (DSP_5) in the distribution (33.68%). On the other hand the MRA of *Betaproteobacteria* remained consistent throughout the DWTP and DWDS. The decrease in MRA of *Actinobacteria*, where it was relatively high in the DWTP (23.42% in the filter effluent) to a much lower level in the DWDS (4.24% at DSP_5) was also reported before.

Despite knowledge of the abundance of each OTU in the DWTP and DWDS, it is still necessary to effectively distinguish whether the abundance measured is based on intact or non-intact cells. The 16S rRNA profiling used for the identification of the OTUs was based on V4-hypervariable region of the 16S rRNA gene, which was isolated and amplified from samples containing total DNA. Apart from the DNA obtained from living cells, DNA released into the bulk water by microbial cells killed during disinfection may also be present in samples. Abundances linked to non-viable DNA is irrelevant as there is no public health risk associated with it and the viability of organisms needs to be investigated. Therefore, flow cytometry was performed on the DWDS samples collected during the last 5 months (September 2016-January 2017) of sampling. A clear decrease in both TCC and ICC was observed between the source water and the various treatment steps. This is expected

as the source water is subjected to coagulation, flocculation, sedimentation, filtration and disinfection. At the ClAm sampling site, an increase in both cell counts was observed, followed by a decrease upon the addition of chloramine to the water. The levels remained low until it started to pick up in the final sampling site. It is meaningful to include viable studies when determining the dynamics of the microbiome of drinking water in order to ensure that correct inferences are made based on 16S profiling data (Proctor and Hammes, 2015).

4.5 SUMMARY

The microbial community inhabiting the drinking water distribution system is strongly shaped by spatial groupings rather than temporal groupings. This was expected as the treatment processes such as filtration and disinfection can have a significant impact on the microbial community along with other physiochemical and environmental factors. The fact that the community present on the sand filter media did not shape the microbiome of the DWDS demonstrated that treatment and distribution systems could differ markedly from each other and that a universal model to predict the microbial community of the water supplied to the consumer would be difficult to achieve. The current study suggests that it is important to study the microbial ecology of individual treatment and distribution systems in order to develop appropriate measures to manage the microbial quality of drinking water in such a system.

CHAPTER 5: POPULATION DYNAMICS OF OTHER DRINKING WATER TREATMENT, DISTRIBUTION AND RETICULATION SYSTEMS

5.1 INTRODUCTION

Drinking water utilities and municipalities face several challenges when providing water to consumers as water with a high microbial numbers can cause major challenges. From the drinking water treatment plant, the water travels through the distribution system and the reticulation system and ultimately to the household tap. At each step of the way, the microbial community composition can be altered by various chemical and physical parameters that have an impact on the microbial community. In order to manage these changes, the microbial ecology within the distribution systems needs to be examined and understood.

The main focus of the current project was to understand the microbial ecology, community diversity as well as the main drivers for change in the microbial communities of drinking water distribution systems. These distribution systems typically represent a complex network of multiple interconnected pipelines and reservoirs with multiple disinfection strategies applied to maintain the microbial quality of the treated water. Such systems are representative of the majority of water utilities that supply water to urban municipalities. Therefore, studies focusing on these systems may provide valuable information related to the microbial ecology of bulk water supply systems. Issues that have not yet been with regards to these systems include the effect of different source waters, treatment procedures and the municipal reticulation systems and their impact on the bacterial community of the drinking water supplied to the consumer.

The focus of this phase of the study was to specifically investigate whether the molecular approach utilised, to understand microbial ecology, would be applicable and beneficial in smaller South African treatment and distribution systems, which utilise different treatment technologies and water sources. Furthermore, the impact of the drinking water reticulation system, characterized by smaller diameter pipes, on the bacterial community in the water was also investigated. For this part of the study the microbial ecology and community dynamics of three additional systems was investigated. Here, a small system supplying chlorinated surface water, a small system supplying groundwater after limited treatment and a municipal reticulation system supplying water obtained from a large water utility were studied.

5.2 MATERIALS AND METHODS

5.2.1 Sampling sites

The impact of source, treatment and reticulation of three different systems, operated under different conditions, are reported on in this chapter.

System 1 – Surface water using an alternative treatment process

The first system was a treatment plant and corresponding distribution system operated by one of the smaller water utilities. This system obtained water directly from a nearby dam and used an alternative treatment regime to the system studied in Chapter 4. Briefly, pre-treatment of the surface water includes flocculation and sedimentation as well as a pre-chlorination step. The pre-treated water is then passed through a combined dissolved air flotation and rapid sand filtration step, which is followed by a final chlorination step before the water is distributed through a 65km pipeline to the municipality and other bulk water users. Five sampling points were included in this study and monthly samples were collected over one year. The source water was sampled directly at the point of extraction (A1), after pre-treatment with chlorine, before filtration (A2), after filtration (A3), final chlorination (A4) and lastly, at the end of the distribution line (A5) (Figure 5.1).

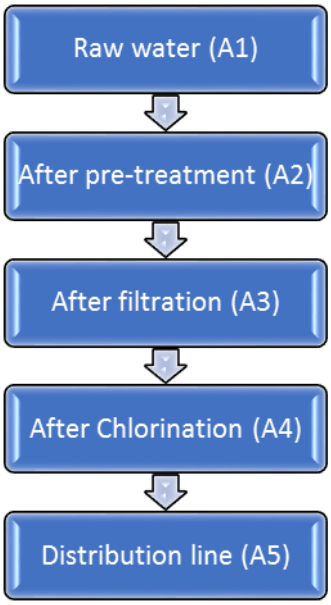


Figure 5.1 Schematic layout of system 1 using surface water and alternative treatment processes.

System 2 – Groundwater system

The second system supplies groundwater, after limited treatment via a reservoir, to an urban community. The water originates from two dolomite aquifers that are separated by a syenite dyke. The groundwater is only chlorinated, without any further treatment such as filtration. The four sampling points included the two springs (i.e. the lower spring (F1) and upper spring (F2)), a reservoir (F3) receiving the chlorinated drinking water from

both groundwater sources and a tap within a dwelling (F4), representing the reticulation section of the distribution system (Figure 5.2). Ten sets of samples were collected on a monthly basis over a period of one year.

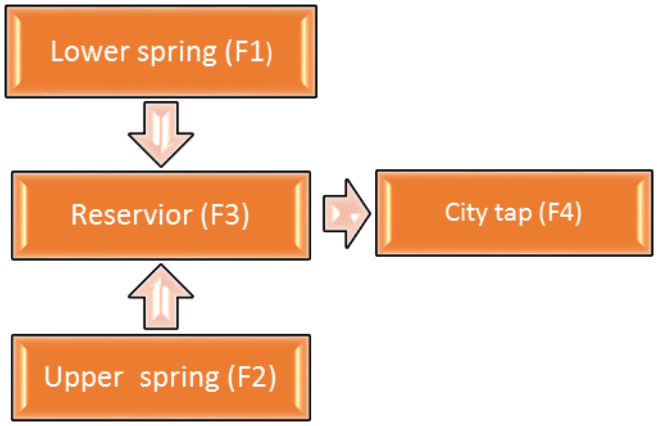


Figure 5.2 Schematic layout of the groundwater system of system 2.

System 3 – Distribution and reticulation systems

The third part of this study is a system that includes a distribution and reticulation system operated by a local municipality, where the main focus is to understand the effects on the microbial community associated with the transition from the large bulk distribution system to the reticulation system. Samples were collected based on the availability of sampling points, and accessibility to the pipeline. During this study 8 points were sampled on a monthly basis over a period of 6 months. Sampling points covered part of the large bulk water distribution system operated by the water utility as well as reservoirs and the reticulation system operated by the municipality. Here, the drinking water receives secondary disinfection (chloramination) at the initial sampling points (CA) but no additional treatment is applied as the water moves through the distribution and reticulation system and to the end-users at the tap.

After the initial samples taken before and after chloramination (CA), samples were collected at the main reservoir (R2 or R_2), which forms part of the bulk water distribution system operated by the water utility. The water then flows into a local reservoir (LR2) under the jurisdiction of the local municipality from where it was distributed to a consumer tap (T2) where another sample was collected. The overall flow can be seen in Figure 5.3.

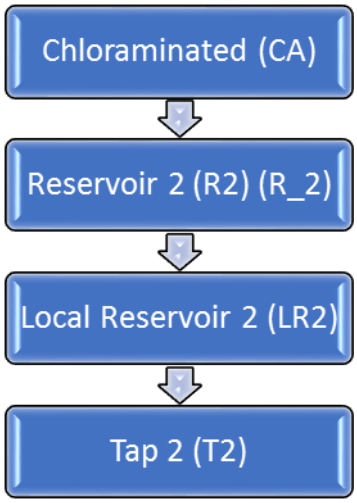


Figure 5.3 Schematic layout of the distribution and reticulation system.

5.2.2 Sample preparation, processing

Samples were collected and processed as described in Section 4.2.2. In the case of the raw groundwater samples (F1 and F2), up to 16 litres of water were concentrated to obtain enough biomass for further analysis.

5.2.3 Flow cytometry

Flow cytometry was performed according to the methodology described in Section 4.2.3.

5.2.4 DNA extraction

DNA was extracted from the biomass collected on the filters as was described in Section 3.2.2.

5.2.5 16S rRNA gene amplification and Illumina MiSeq sequencing

The same procedure as described in Section 3.2.2 was followed and samples were again sequenced at the University of Michigan Medical School (Ann Arbor, United States of America), using a pair-ended sequencing protocol as described by Kozich and colleagues (2013).

5.2.6 Sequence processing and data analysis

MOTHUR sequence processing

The sequence and data processing were conducted using MOTHUR according to the procedure described in Section 3.2.3.

Alpha and beta diversity analysis

Alpha and beta diversity indexes and the comparison thereof were also performed as described in Section 3.2.3.

Statistical analysis

For statistical analyses MOTHUR and R were used as described in Section 3.2.3.

5.3 RESULTS

5.3.1 System 1 – Surface water using an alternative treatment process

Sequence data

Due to failed sequencing runs, two samples from sampling point A5 (months 3 and 12) had to be removed prior to sequence processing. Removal of these sequences did not affect the data set as 96% of the data was still represented. A total of 2 273 985 raw sequence reads were generated from the sampling points, over the period of 12 months. After the removal of sequences with ambiguous bases and anything longer than 275bp (set by MOTHUR), the number of sequences was reduced to 1 771 509. Further processing and filtration resulted in the retention of 1 720 907 sequences with the average number of sequences per sample were 24 940 ± 18 101 (M ± SD). A total of 9 325 Operational Taxonomic Units (OTUs) were observed from all the sequence data collected. For further alpha and beta diversity analyses, subsampling was performed a 1000 times at the minimum number of sequences observe in a sample (n = 279).

Community membership

The overall mean relative abundances (MRA) were calculated for the five sample points averaged over the period of 12 months. Figure 5.4 gives a representation of the MRA at phylum level. A diverse bacterial community was observed across all sampling points. *Proteobacteria* showed the highest MRA across all sampling points (i.e. A1, M ± SD = 21 ± 3%; A2, M ± SD = 36 ± 21%; A3, M ± SD = 67 ± 16%; A4, M ± SD = 35 ± 23% and A5, M ± SD = 59 ± 23%). *Cyanobacteria* also showed high MRAs in samples A2, A4 and A5 (A2, M ± SD = 27 ± 19%; A4, M ± SD = 23 ± 25% and A5, M ± SD = 16 ± 20%). *Bacteroidetes* had a high MRA at sample point A1 with *M* ± *SD* of 41 ± 12% and fairly high level at all other sample points (i.e. A2, M ± SD = 14 ± 8% was measured, followed by A3 with M ± SD = 14 ± 8%, A4 with M ± SD = 12 ± 10% and A5 with M ± SD = 10 ± 8%). A slight decrease in the abundance of *Bacteroidetes* could be seen within the treatment plant. *Verrucomicrobia* had a low mean relative abundance at the start of treatment system (A1) with a MRA of 7 ± 3.6 and almost completely disappeared from the treatment system as seen at point A5 with M ± SD = 0.6 ± 0.7%.

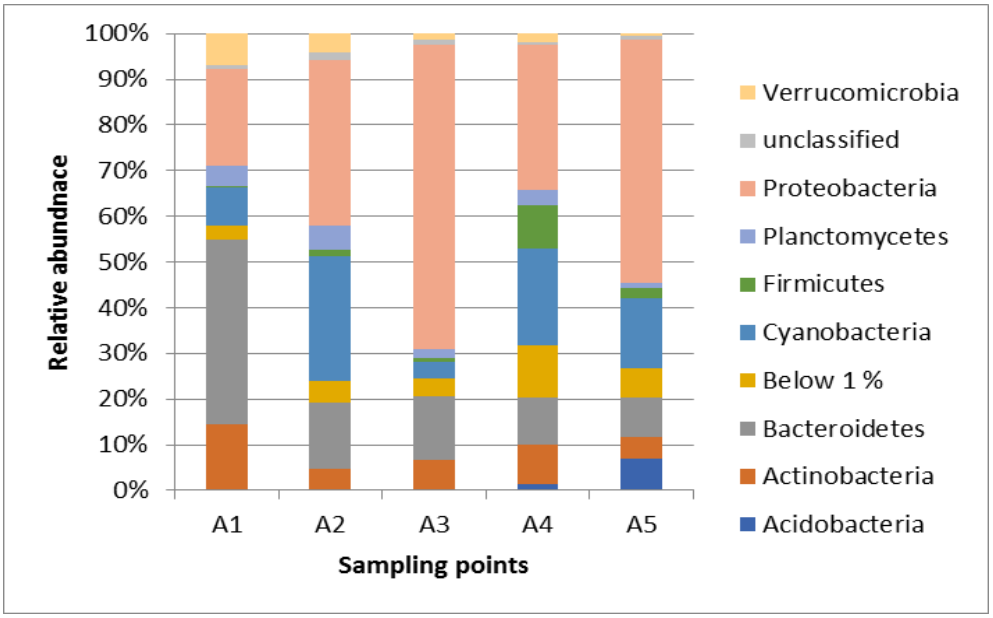


Figure 5.4 Mean relative abundances of the dominant phyla across all five sampling points over a period of 12 months.

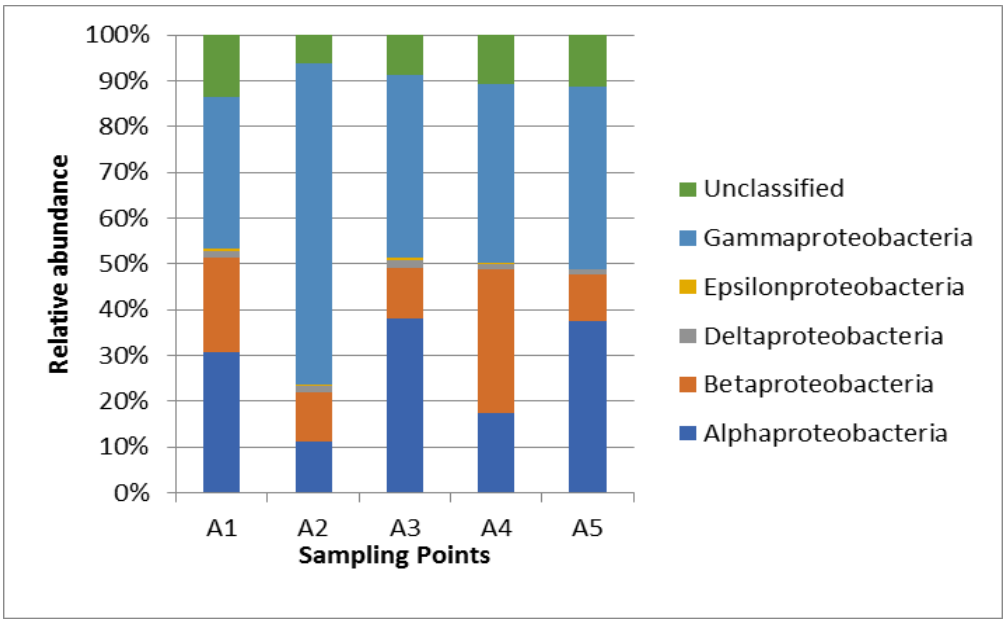


Figure 5.5 Mean relative abundances of the proteobacterial across all five sampling points over a period of 12 months.

Knowing that the *Proteobacteria* had the highest mean relative abundances, a further breakdown of the *Proteobacteria* was done across all sample points. Looking at the MRA of the proteobacterial classes showed that *Gammaproteobacteria* had the highest MRA across the treatment plant (i.e. A1, M ± SD = 16 ± 6%; A2, M ± SD = 47 ± 14%; A3, M ± SD = 21 ± 15%; A4, M ± SD = 22 ± 19% and A5, M ± SD = 22 ± 24%). *Alphaproteobacteria* have the second highest MRA for samples A1, A2, A3 and A5 (i.e. A1, M ± SD = 15 ± 3%; A2, M ± SD = 7 ± 4%; A3, M ± SD = 20 ± 11% and A5 M ± SD = 19 ± 17%), while the *Betaproteobacteria* had the second highest MRA at the point A4 with M ± SD = 17 ± 20% (Figure 5.5).

A heatmap was constructed showing the highly abundant OTUs that constituted >1% of the total sequence abundance (Figure 5.6). Amongst the persistent OTUs, that formed a stable part of the community throughout the system, a member of the genus *Planktothrix* (*Cyanobacteria)* was the most common. Several OTUs were commonly detected during treatment and were observed in the source water as well as in the inflow and outflow of the filter beds, but were removed after the final chlorination step. These included members of the genera *Paludibacter* and *Sediminibacterium* (*Bacteroidetes*) and *Acrobacter* and *Vogesella* (*Proteobacteria*).

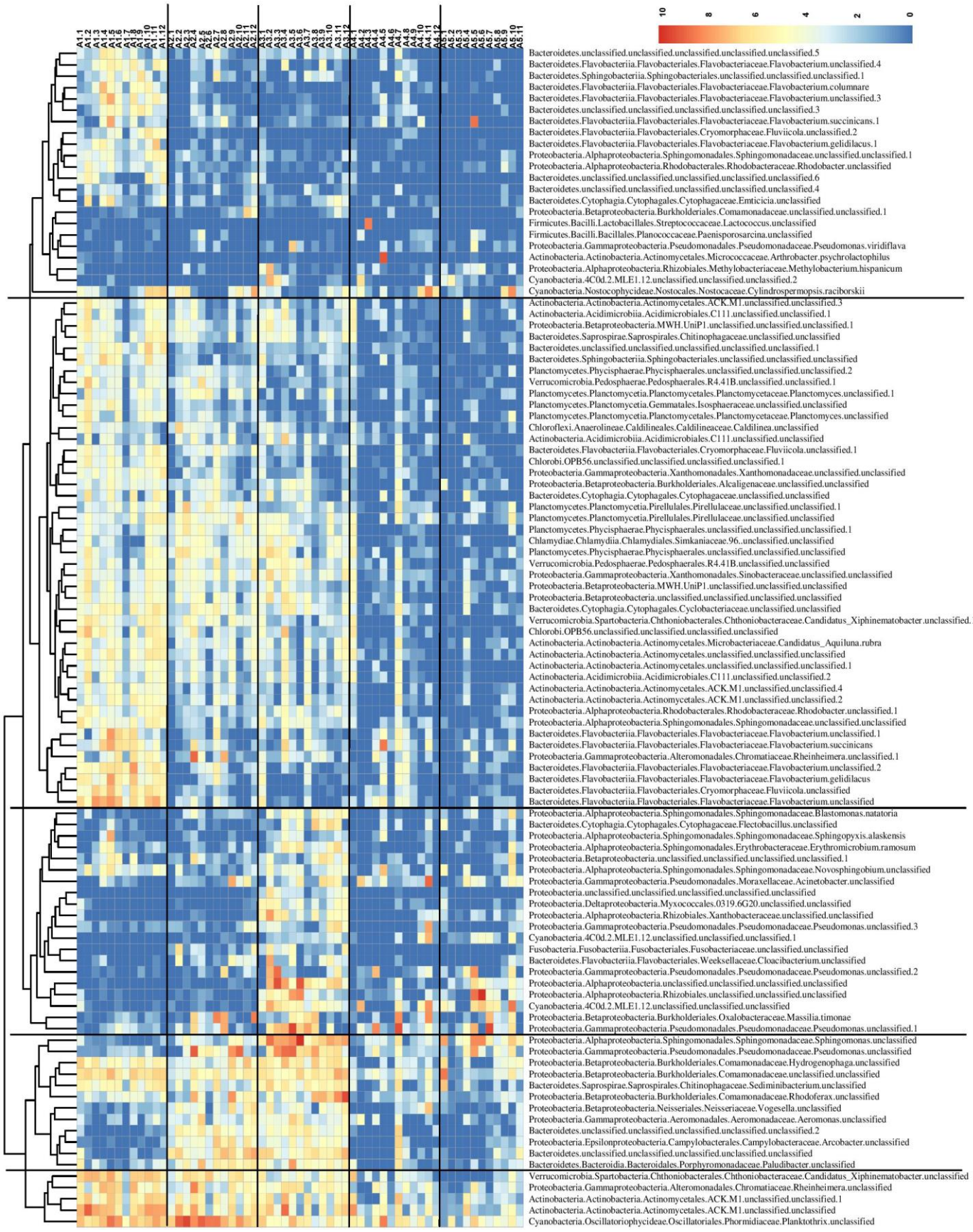


Figure 5.6 Heatmap showing the distribution of the most dominant OTUs in the system using an alternative treatment approach.

Community diversity

Based on the alpha diversity data (Species observed (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J)), significant spatial differences (ANOVA, all $p < 0.05$; Table 5.1, Figure 5.7) were observed for microbial communities across sampling points. However, no significant temporal differences were observed across sample locations. The Tukey's post-hoc HSD test was used to determine between which specific sampling points a significant difference occurred in terms of the overall community diversity characteristics. Differences in the observed species (S_{obs}) were observed between points A4-A1 and A5-A1, Shannon diversity (H') showed differences between points A4-A1 and A5-A1 and Pielou's evenness (J) showed differences between points A2-A1, A4-A1 and A5-A1 when samples were grouped spatially (Tukey's post-hoc HSD test, $p < 0.05$; Table 5.2). Temporal changes in overall community composition across all locations were examined and no significant differences were observed when analysing the data based on monthly or seasonal groupings. This finding was also confirmed with Tukey's post-hoc HSD test showing no significant differences between individual seasons or months (Tables C1-C4).

Table 5.1 ANOVA table and P-values for observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J), evaluating an overall sample points statstical significant effect in community

Observed species (S_{obs})					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	4	11706	2926.54	4.5459	0.00313*
Within the groups	53	34120	643.77		
Total	57				
Shannon diversity (H')					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	4	10.782	2.69542	4.6683	0.002656*
Within the groups	53	30.602	0.57739		
Total	57				
Pielou's evenness (J)					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	4	0.25054	0.062636	3.9261	0.007276*
Within the groups	53	0.84555	0.015954		
Total	57				

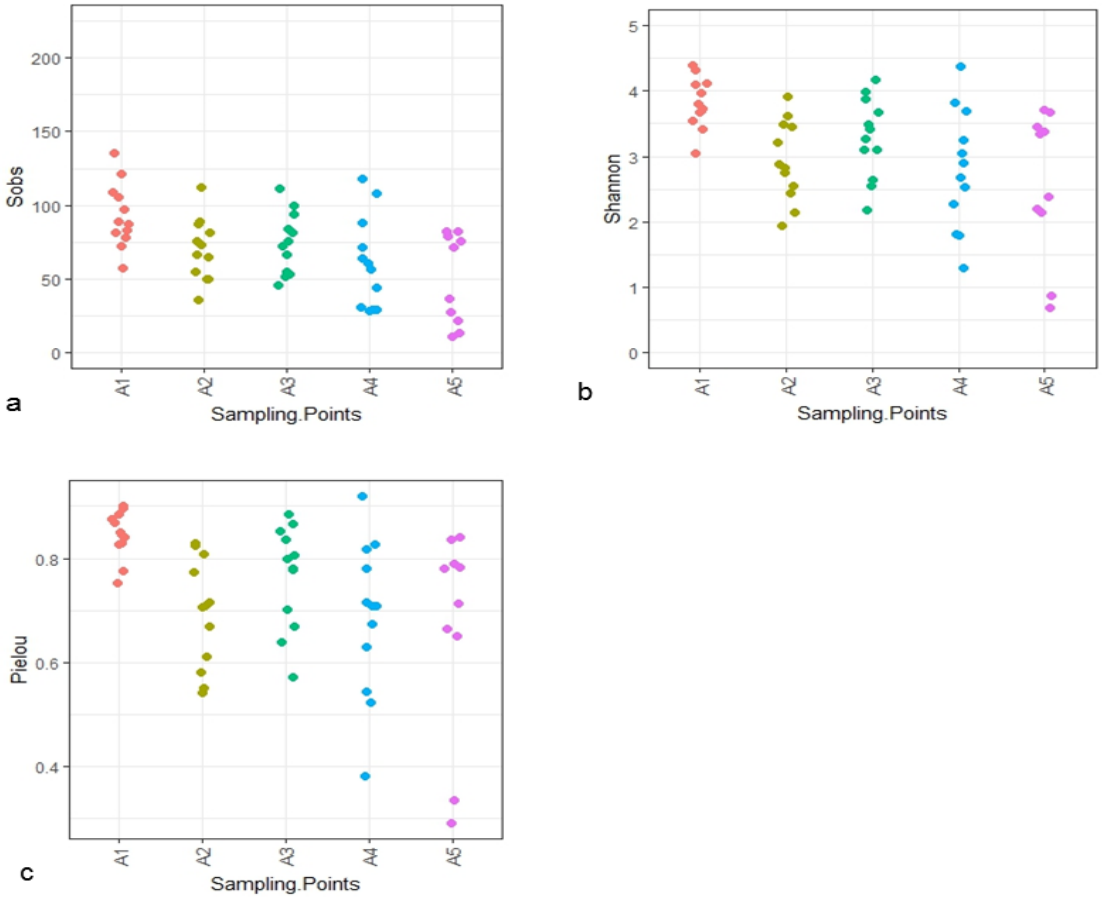


Figure 5.7 Observed species (S_{obs}), Shannon diversity (H') and Pielou's evenness (J) values as determined for samples collected at all five sampling points over a 12 month period.

Table 5.2 Tukey's post-hoc HSD test p-values for observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J), evaluating sampling points statstical significant effect in community

Sample comparisons	Observed species (S_{obs})	Shannon diversity (H')	Pielou's evenness (J)
A2-A1	0.1800181	0.0477151	0.0373675*
A3-A1	0.3692426	0.4343291	0.5336430
A4-A1	0.0221751*	0.0135762*	0.0247604*
A5-A1	0.0019328*	0.0033039*	0.0159016*
A3-A2	0.9935244	0.7905957	0.6380636
A4-A2	0.8967366	0.9891664	0.9998536
A5-A2	0.3648813	0.8106315	0.9906116
A4-A3	0.6842497	0.5006490	0.5377982
A5-A3	0.1840617	0.2082436	0.3946138
A5-A4	0.8631918	0.9684548	0.9979879

The beta diversity data was used to determine the level of dissimilarity between the micobial communities associated with the collected samples. Bray-Curtis and weighted Unifrac distances were used to compare community structure (absence and presence as well as abundance), while Jaccard and unweighted Unifrac distances were used to determine community membership (absence and presence). For the beta diversity measures, values closer to one indicate higher dissimlarity between communities. Summarising the disimilarity distances between each sampling point indicated that the community structure at A4 was the most dissimilar with a average Bray-Curtis value of 0.94 (minimum and maximun values of 0.55 and 0.98, respectively) and an average weighted UniFrac distance of 0.78 (minimum and maximun values of 0.37 and 0.94, respectively). In terms of community membership, the community at A5 was the most disimilar, based the average Jaccard distance of 0.92 (minimum and maximun values of 0.81 and 0.97, respectively) and and average unweighted UniFrac distance of 0.80 (minimum and maximun values of 0.72 and 0.86, respectively).

When performing an AMOVA analysis it was confirmed that all four beta diversity metrics supported significant differences in the community structure and membership when samples were grouped spatially (Table 5.3). When these observed spatial differences were further investigated, by comparing all samples taken during the same month, the AMOVA analysis indicated that significant differences were observed between all the sampling points (Table 5.4) apart from the communities associated with points A4 and A5.

Table 5.3 AMOVA table for distance matrics on Bray-Curtis, unifrac weighted, Jaccard and Unifrac Unweighted between all sampling points (A1-A2-A3-A4-A5). Significant differences between overall locatins are indicated with an *.

Bray-Curtis					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	4.34241	4	1.0856	3.36931	<0.001*
Within	17.0768	53	0.322204		
Total	21.4192	57			
Weighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	4.16429	4	1.04107	5.12926	<0.001*
Within	10.7573	53	0.202968		
Total	14.9216	57			
Jaccard					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	3.58131	4	0.895328	2.61232	<0.001*
Within	18.1648	53	0.342732		
Total	21.7461	57			
Unweighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	3.84258	4	0.960644	3.39569	<0.001*
Within	14.9937	53	0.282901		
Total	18.8363	57			

Table 5.4 AMOVA showing distance matrices on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac, showing where the significant difference are (shown with a *)

Sample comparisons	Bray-Curtis	Weighted UniFrac	Jaccard	Unweighted UniFrac
A1-A2	<0.001*	<0.001*	<0.001*	<0.001*
A1-A3	<0.001*	<0.001*	<0.001*	<0.001*
A1-A4	<0.001*	<0.001*	<0.001*	<0.001*
A1-A5	<0.001*	<0.001*	<0.001*	<0.001*
A2-A3	<0.001*	<0.001*	<0.001*	<0.001*
A2-A4	<0.001*	<0.001*	0.001*	<0.001*
A2-A5	<0.001*	<0.001*	<0.001*	<0.001*
A3-A4	<0.001*	<0.001*	<0.001*	<0.001*
A3-A5	0.009*	0.0258	<0.001*	<0.001*
A4-A5	0.251	0.161	0.92	0.535

Comparing sampling points on a monthly or seasonal basis, based on all beta diversity metrics, also indicated significant differences between sampling points as opposed to the alpha diversity measures (Table 5.5, 5.6 and C5). When investigating these temporal changes it was observed that more significant changes occurred in community structure (Bray-Curtis and weighted UniFrac) than for community membership (Jaccard and unweighted UniFrac) (Table 5.6). Significant differences based on monthly comparisons were mainly observed between the months in which sampling was done at least three months apart (Table C6). Furthermore, the PCoA ordination (Figure 5.8) indicated that samples grouped primarily according to sample points (spatial grouping) and that seasonal changes played less of an important role. A few outliers were observed for all the sampling points but samples from points A4 and A5 could be grouped together.

Table 5.5 AMOVA showing distance matrices on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all seasons in relation to one another (Autumn-Spring-Summer-Winter), showing where the significant difference were found (shown with a *)

Bray-Curtis					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	1.98043	3	0.660144	1.83385	<0.001*
Within	19.4388	54	0.359977		
Total	21.4192	57			
Jaccard					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	1.78125	3	0.593751	1.60595	<0.001*
Within	19.9649	54	0.36972		
Total	21.7461	57			
Unweighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	1.48521	3	0.495071	1.54076	0.002*
Within	17.3511	54	0.321317		
Total	18.8363	57			
Weighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	1.3697	3	0.456566	1.81927	0.002*
Within	13.5519	54	0.250961		
Total	14.9216	57			

Table 5.6 AMOVA table for distance matrics on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all sampling points, showing significant differences to one another (shown with a *)

Sample comparisons	**Bray-Curtis**	**Weighted UniFrac**	**Jaccard**	**Unweighted UniFrac**
Autumn-Spring	<0.001*	0.002*	0.006*	0.008*
Autumn-Summer	0.029*	0.068	0.24	0.169
Autumn-Winter	0.045*	0.014*	0.113	0.099
Spring-Summer	0.027*	0.019*	0.114	0.044*
Spring-Winter	0.018*	0.011*	0.028*	0.029*
Summer-Winter	0.003*	0.013*	0.013*	0.035*

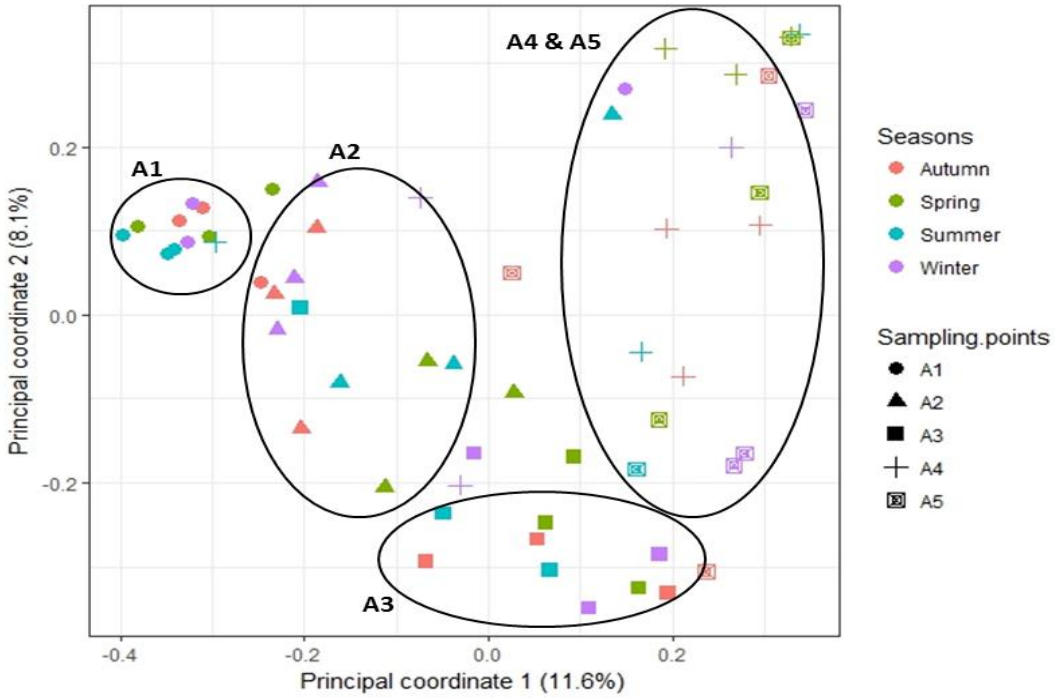


Figure 5.8 PCoA plot showing the community relation of all five sampling points over 12 months with corresponding seasonal changes.

Flow cytometry

From the flow cytometry data (Figure 5.9), a distinct decrease in cell numbers could be observed, from the start of the treatment to the point where it was measured in the distribution system. At all sampling points, except A1, the intact cell counts were higher than the damaged cell counts.

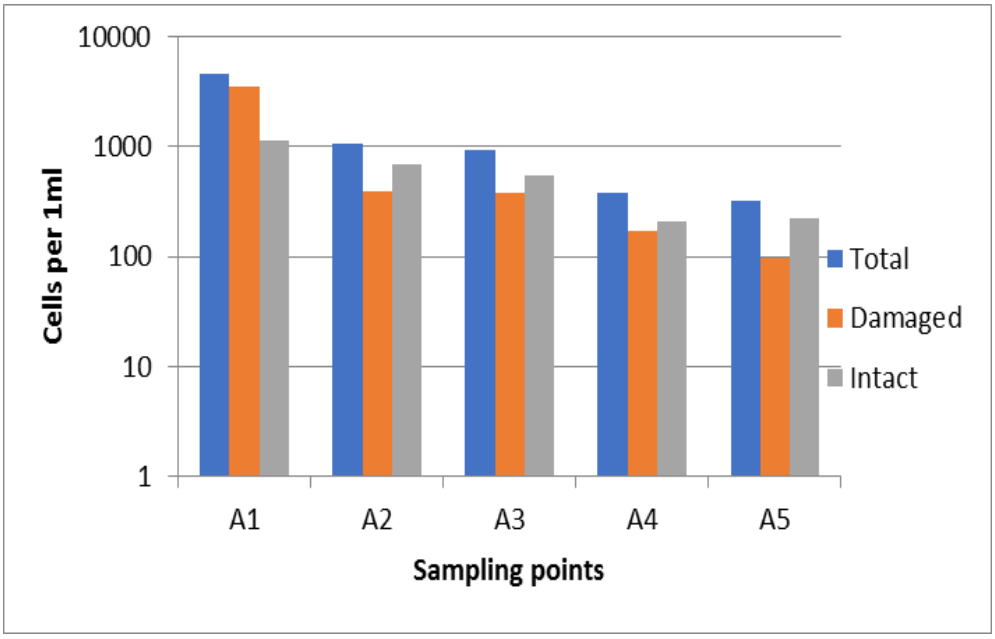


Figure 5.9 Flow cytometry results averaged over a five month period for all five sampling points, representing total cell count, dead cell count and alive cell count in 1 ml of water.

5.3.2 System 2 – Groundwater system

Sequence data

The water supply system that delivers groundwater to consumers was sampled for a period of 11 months (March 2016 to January 2017). However, sampling did not take place during July 2016 and no sample was also collected for point F3 during January 2017 due to problems to obtain access to the sampling sites. In addition, due to failed sequence runs, data from F1 (months 6 and 10), F4 (month 8) and F2 (month 3) were excluded from further sequence analysis. The remaining sequences still represented far more than 90% of the complete data set. Following Illumina MiSeq sequencing of the V4-hypervariable region of the 16S rRNA gene, processing pipeline 1 020 406 raw sequence reads were observed. Sequence data was then processed using MOTHUR and after removal of sequences with ambiguous bases and anything longer then 275bp, the number of sequences were reduced to 779 249. After further processing and quality filtering of the data, 766 162 sequences were retained. The average number of sequences per sample is 21 282 ± 20 094 (M ± SD) and resulting in the identification of 8 013 OTUs after a 97% similarity threshold. For further alpha and beta diversity analyses, subsampling was performed a 1000 times at the minimum number of sequences observe in a sample (n = 279).

Community membership

Overall, a diverse bacterial community was observed (Figure 5.10). *Proteobacteria* was observed to be the most dominant phyla with the highest mean relative abundance (MRA) across all sampling points (i.e. F1, M ± SD = 49 ± 12%; F2, M ± SD = 50 ± 19%; F3, M ± SD = 42 ± 21% and F4, M ± SD = 59 ± 24%. The *Actinobacteria* was present at low MRAs for sampling points F1, F3 and F4 (F1, M ± SD = 5 ± 5%; F2, M ± SD = 4 ± 7% and F4, M ± SD = 5 ± 19%), however it had a MRA at sampling point F3 with M ± SD of 37 ± 4%. *Bacteroidetes* had the second highest MRA for sample points F1, F2 and F4 (F1, M ± SD = 19 ± 13%; F2, M ± SD = 13 ± 11% and F4, M ± SD = 11 ± 4%. Here, it was clear from the data that point F3 (a reservoir) had a community membership that differed from the other sampling points.

Based on the observation that the *Proteobacteria* dominated the system, the distribution of this group of bacteria was further investigated (Figure 5.11). The dominant proteobacterial groups were *Alpha*-, *Beta*- and *Gammaproteobacteria* in the system. *Gammaproteobacteria* had the highest MRAs for sampling points F1, F2 and F4 (F1, M ± SD = 24 ± 29%; F2, M ± SD = 26 ± 22% and F4, M ± SD = 26 ± 27%. In contrast the *Alphaproteobacteria* had the highest MRA at point F3 (M ± SD = 52 ± 25%).

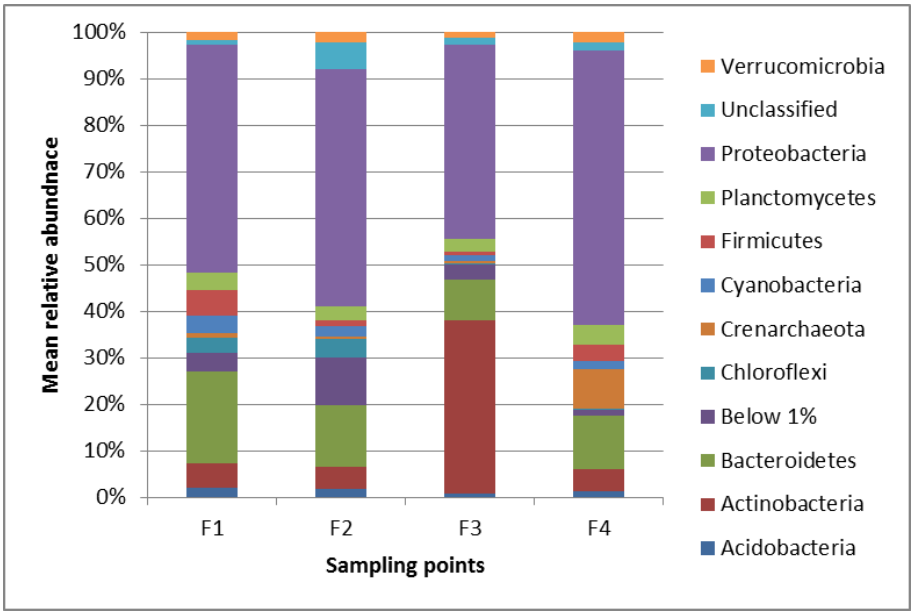


Figure 5.10 Mean relative abundances of the dominant phyla across all four sampling points over a period of 11 months.

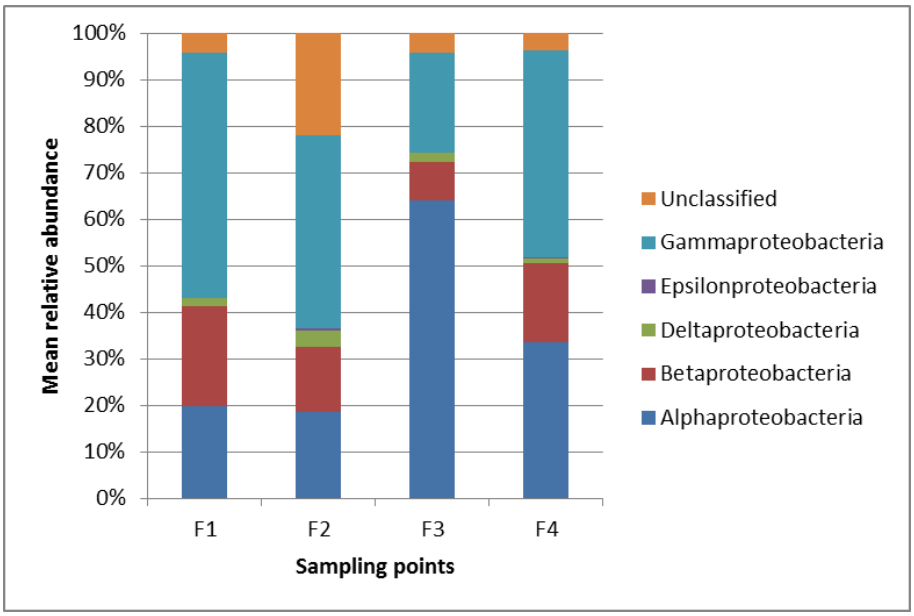


Figure 5.11 Mean relative abundances of the proteobacterial classes across all four sampling points over a period of 11 months.

As with the previous system, a heatmap (Figure 5.12) was constructed to display all the highly abundant OTUs that constituted >1% of the total sequence abundance. Several OTUs were present throughout the system and included members of the genera such as *Pseudomonas, Actinobacter and Flavobacterium.* OTUs that were dominant in the reservoir samples and reticulation system included members of genera such as *Nitrosomonas* and *Hyphomicrobium*.

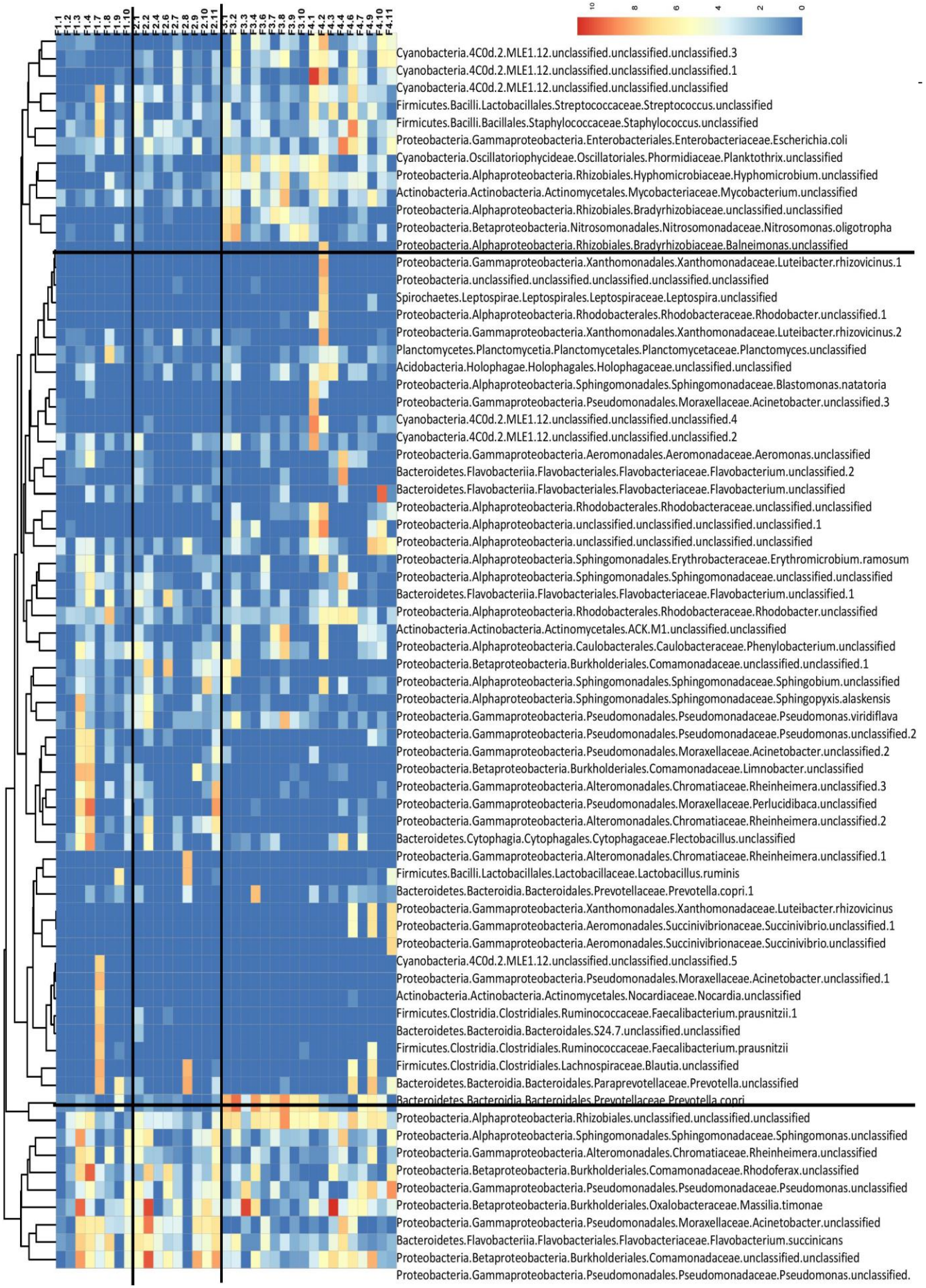


Figure 5.12 Heatmap showing the distribution of the most dominant OTUs in a groundwater.

Community diversity

Statistical analysis of the alpha diversity data showed no significant differences between communities in terms of the species observed (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J). Based on the alpha diversity measures at each sampling point there were no significant spatial differences between the communities associated with the different sampling points (ANOVA, all $p > 0.05$, Table C7) or when individual samplings points were compared with each other (Table C8). There were also no significant temporal differences between communities within the same sampling point when the data was analysed on a seasonal or month to month basis (Table C9 and C11). Furthermore, a Tukey's post-hoc HSD test was done to ensure no significant differences between individual months or seasons were observed (Table C10 and C12).

Although the statistical analysis revealed no significant differences in alpha diversity measures when samples were grouped either spatially or temporally, there were noticeable differences observed within each sampling point (Figure 5.13). Observed species (S_{obs}) showed variable richness across all samples, specifically for point F2, where richness was highly variable across the duration of the study in comparison to sampling points F1, F3 and F4. Shannon diversity index (H') showed a range of values at each point, with F2 again being more variable. Pielou's evenness (J) with only Point F1 demonstrating limited variation.

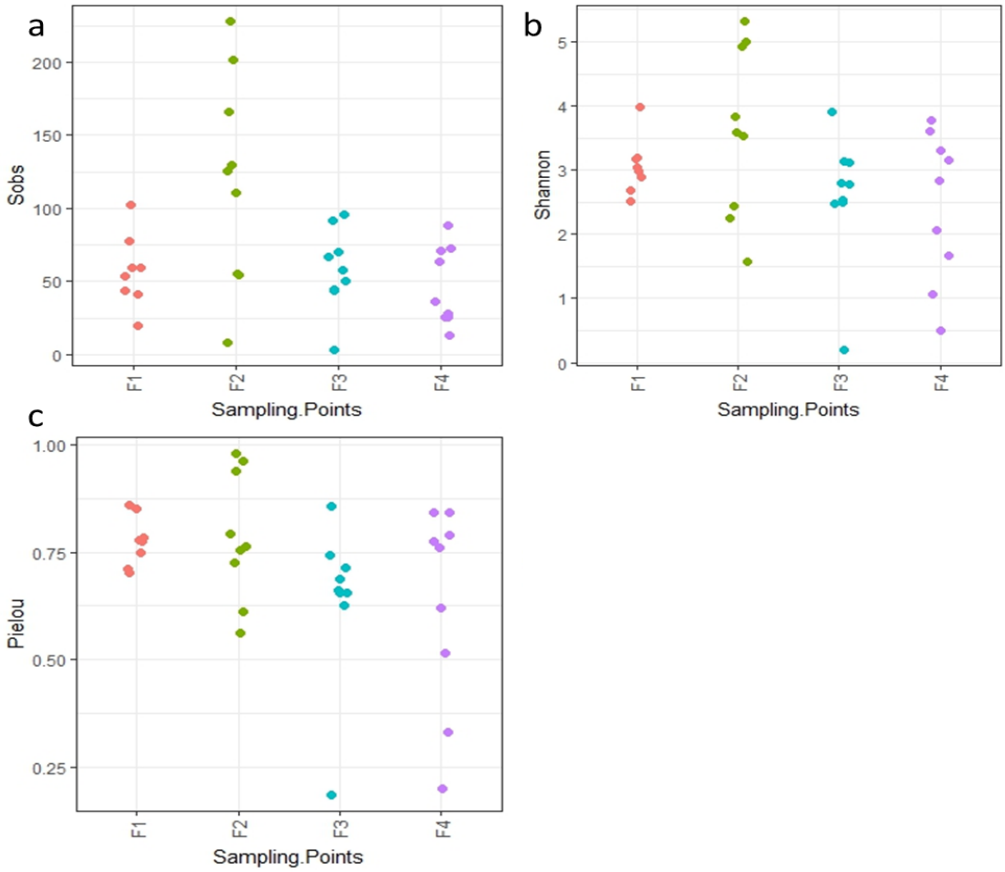


Figure 5.13 Observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J) values for all communities sampled across all four sampling points over a11 month period.

Based on the OTU abundance data, differences between the community structure and membership of individual communities were assessed (beta diversity). Based on all four of the metrics, it was clear that significant differences existed between that the communities associated with the different sampling points (Table 5.7). When an AMOVA test was performed, to determine between which points these significant differences occurred (Table 5.8), it was clear that, F2 differed significantly from all other sampling points. It was only based on community structure (Bray-Curtis and weighted UniFrac), that no significant differences were observed between F2 and F1. In terms of community membership, F1 (lower spring source) and F3 (reservoir) showed no significant differences but the structural differences were found to be significant. The opposite situation was true in that F1 (lower spring source) and F4 (tap) were similar in terms of community structure but differed in terms of community membership. Statistical analysis showed that no significant temporal differences (Table C13 and C14) were observed when the data was compared on a monthly or seasonal basis.

Table 5.7 AMOVA table for distance matrics on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all sampling points (F1-F2-F3-F4), showing significant differences to one another (shown with a *)

Bray-Curtis					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.32248	3	0.77416	1.87301	<0.001*
Within	12.813	31	0.413323		
Total	15.1355	34			
Jaccard					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.09957	3	0.699856	1.72991	<0.001*
Within	12.5414	31	0.404561		
Total	14.641	34			
Unweighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.21865	3	0.73955	2.12722	<0.001*
Within	10.7775	31	0.34766		
Total	12.9961	34			
Weighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.40578	3	0.801928	2.81426	<0.001*
Within	8.83349	31	0.284951		
Total	11.2393	34			

Table 5.8 AMOVA test showing distance matrices on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all sampling points in relation to one another, showing where the significant difference were found (shown with a *)

Sample comparisons	Bray-Curtis	Weighted UniFrac	Jaccard	Unweighted UniFrac
F1-F2	0.139	0.051	0.006*	0.015*
F1-F3	<0.001*	0.002*	0.08	0.068
F1-F4	0.423	0.316	0.023*	0.014*
F2-F3	<0.001*	<0.001*	<0.001*	<0.001*
F2-F4	<0.001*	<0.001*	<0.001*	<0.001*
F3-F4	0.005*	<0.001*	0.01	0.004*

The PCoA plot based on the data above (Figure 5.14) showed that the data could be clustered on two broad groups. The first was a cluster which grouped most of the samples from F1, F3 and F4, with the most of the communities for F3 (reservoir) a bit more separated from the rest of the samples in that cluster. The second was a clustering of most of the samples representing F2. A few outliers associated with autumn were also observed. This data showed that the communities within the system was mainly impacted by the community present in the main source (F1) and that seasonality had a limited impact.

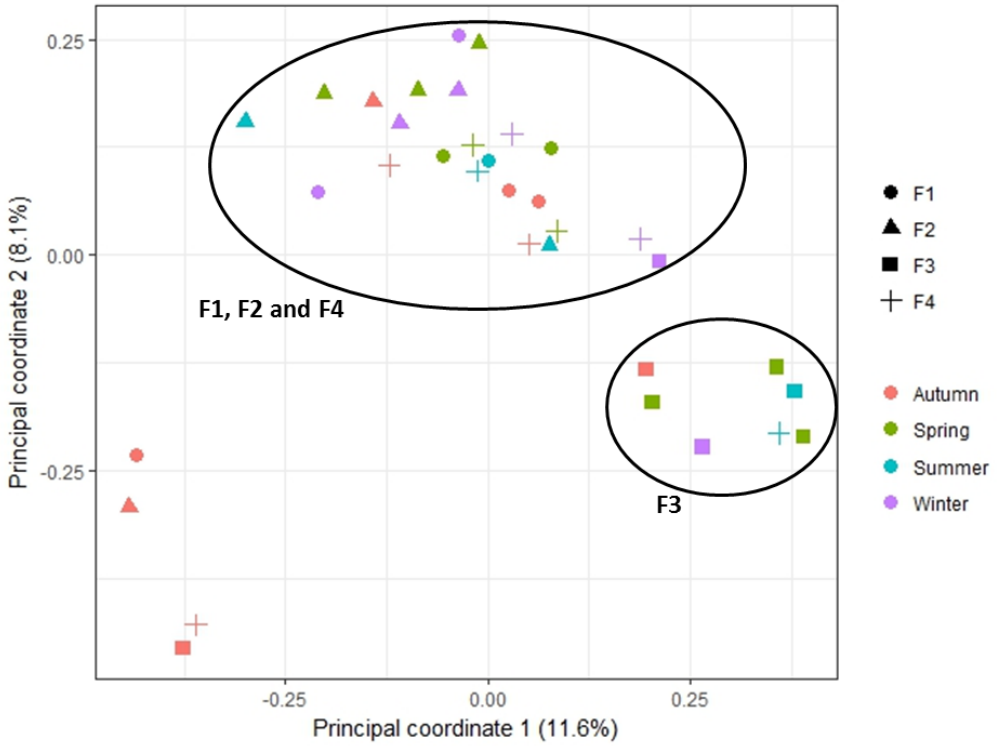


Figure 5.14 PCoA plot showing the community relation of all for samling points over 11 months with corrosponding seasonal changes.

Flow cytometry

The average flow cytometry results, as measured over a period of 5 months for all four sampling points are presented in Figure 5.15. Total cell counts were low in both source water samples (F1 and F2), while the level of bacteria increased nearly by two logs within the reservoir (F3). Thereafter, a one log reduction was seen between the reservoir and the bacterial level as measured at the tap of the consumer (F4).

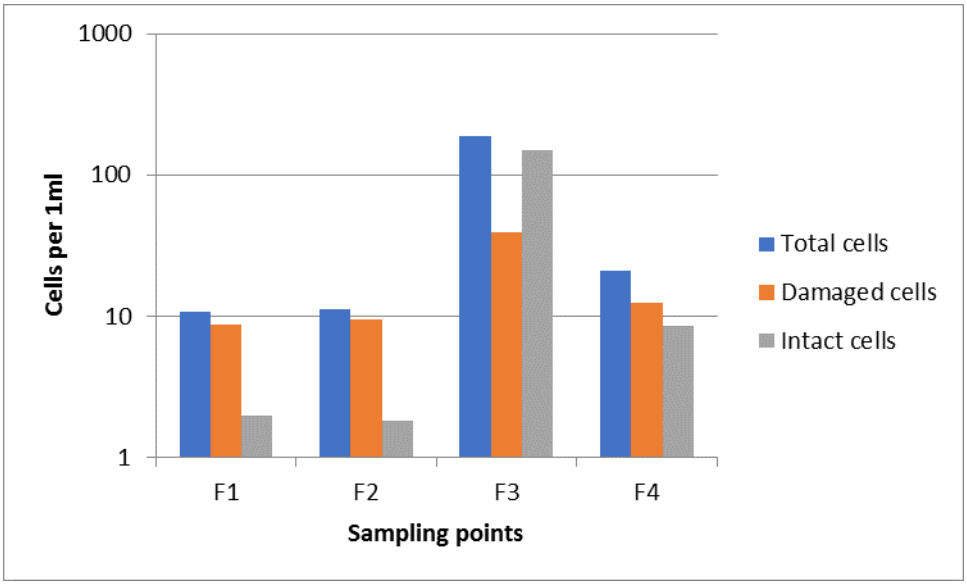


Figure 5.15 Flow cytometry results averaged over a five month period for all four sampling points, representing total cell count, dead cell count and alive cell count in 1 ml of water.

5.3.3 System 3 – Distribution and reticulation system

Sequence data

The sampling of the combined distribution and reticulation system was done over a period of 6 months. Due to initial sequence failures, a few samples had to be removed from the original dataset. These included sampling point CA (months 2, 4 and 6) as well as R2 (month 3) and LR2 (month 3). Removal of these sequences was acceptable as the remaining sequences still represented more than 90% of the original data. Following Illumina MiSeq sequencing, a total of sequences 1 113 542 were observed. Processing of the data with MOTHUR resulted in the removal of sequences with ambiguous bases and anything longer than 275bp and the number of sequences was reduced to 1 1054 533 ± 12203 (M ± SD). The average number of sequences per sample was 18 830 and a total of 4 301 Operational taxonomic units (OTUs) were identified. To create an evenly spread dataset for the alpha and beta statistical analyses, sub-sampling was done a 1000 times (*n* = 112).

Community membership

The overall mean relative abundances (MRA) were calculated at the phylum level. The data indicated a diverse bacterial community spread across all sampling points. *Proteobacteria* was observed to have the highest MRA in all sampling points (i.e. CA, M ± SD = 25 ± 4%; R2, M ± SD = 6 ± 6%; LR2, M ± SD = 2 ± 2% and T2, M ± SD = 7 ± 5%) (Figure 5. 16). At the same time a decrease in the levels of *Bacteroidetes* was observed, from 25% to 7%.

Based on the fact that the *Proteobacteria* had the highest MRA, a further breakdown of *Proteobacteria* across all sample points was performed. Looking at the MRAs at class level, the *Alpha-, Beta-* and *Gammaproteobacteria* dominated the system. *Alphaproteobacteria* was the most abundant across all sampling points (i.e. MRA: CA, M ± SD = 34 ± 21%; R2, 45 15%; LR2, 39 ± 20% and T2, 34% ± 18%), with *Betaproteobacteria* having the second highest *MRA* (i.e. CA, M ± SD = 1 ± 3%; R2, M ± SD = 18 ± 1%; LR2, M ± SD = 22 ± 16% and T2, M ± SD = 21 ± 8%) (Figure 5.17).

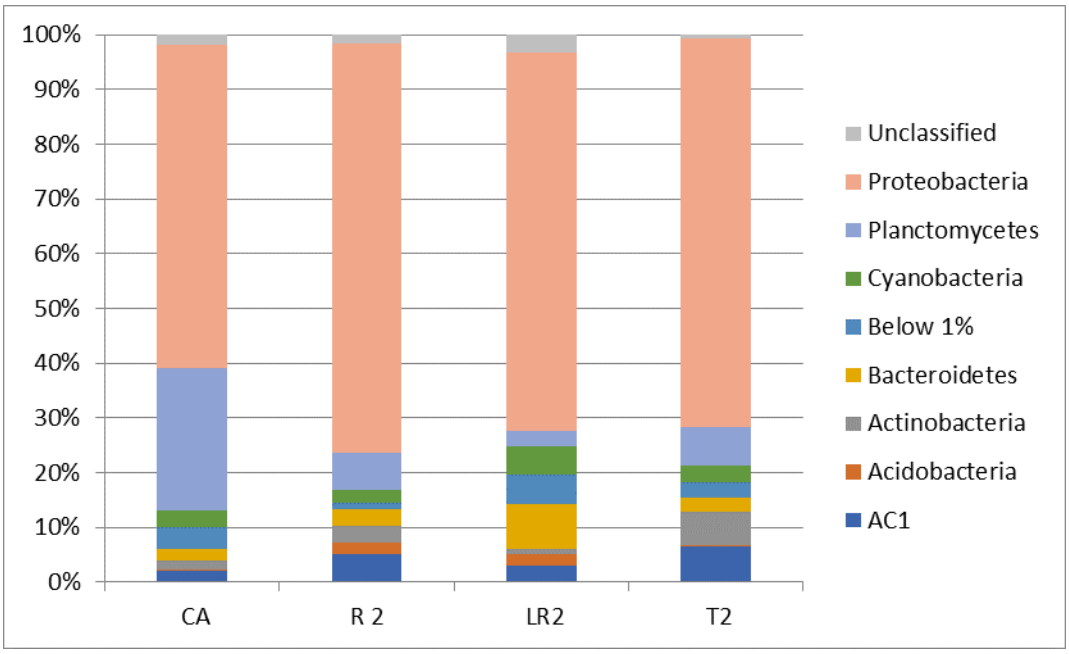


Figure 5.16 Mean relative abundances of the dominant phyla across all sampling points over a period 6 of months.

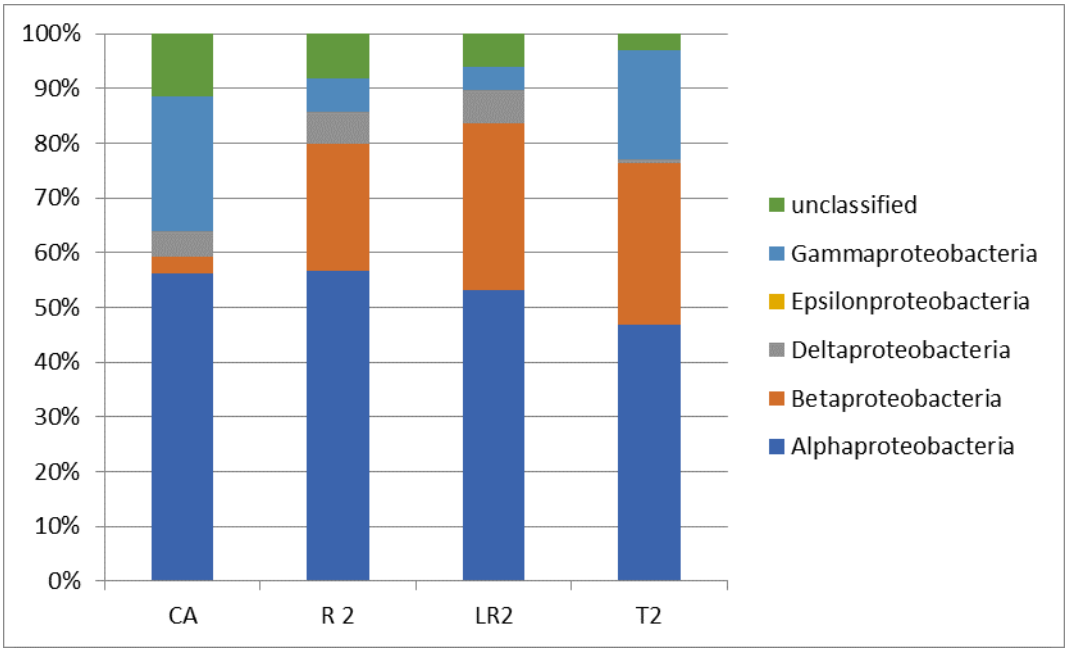


Figure 5.17 Mean relative abundances of the proteobacterial classes across all sampling points over a period of 6 months.

The OTUs with an abundance of at least 1% in the combined dataset were plotted as a heatmap (as done for the previous 2 systems) according to the four sampling points (Figure 5. 18). Several bacteria such as members of the genera *Nitrosomonas, Planctomyces* and *Sphingomonas* were present at high abundances throughout the system, whereas other OTUs were only seen at some of the sampling points. None of the OTUs detected at the final tap were recent introductions and the data suggested that the community, as measured at the tap, was a product of the communities that have previously developed in the larger distribution system.

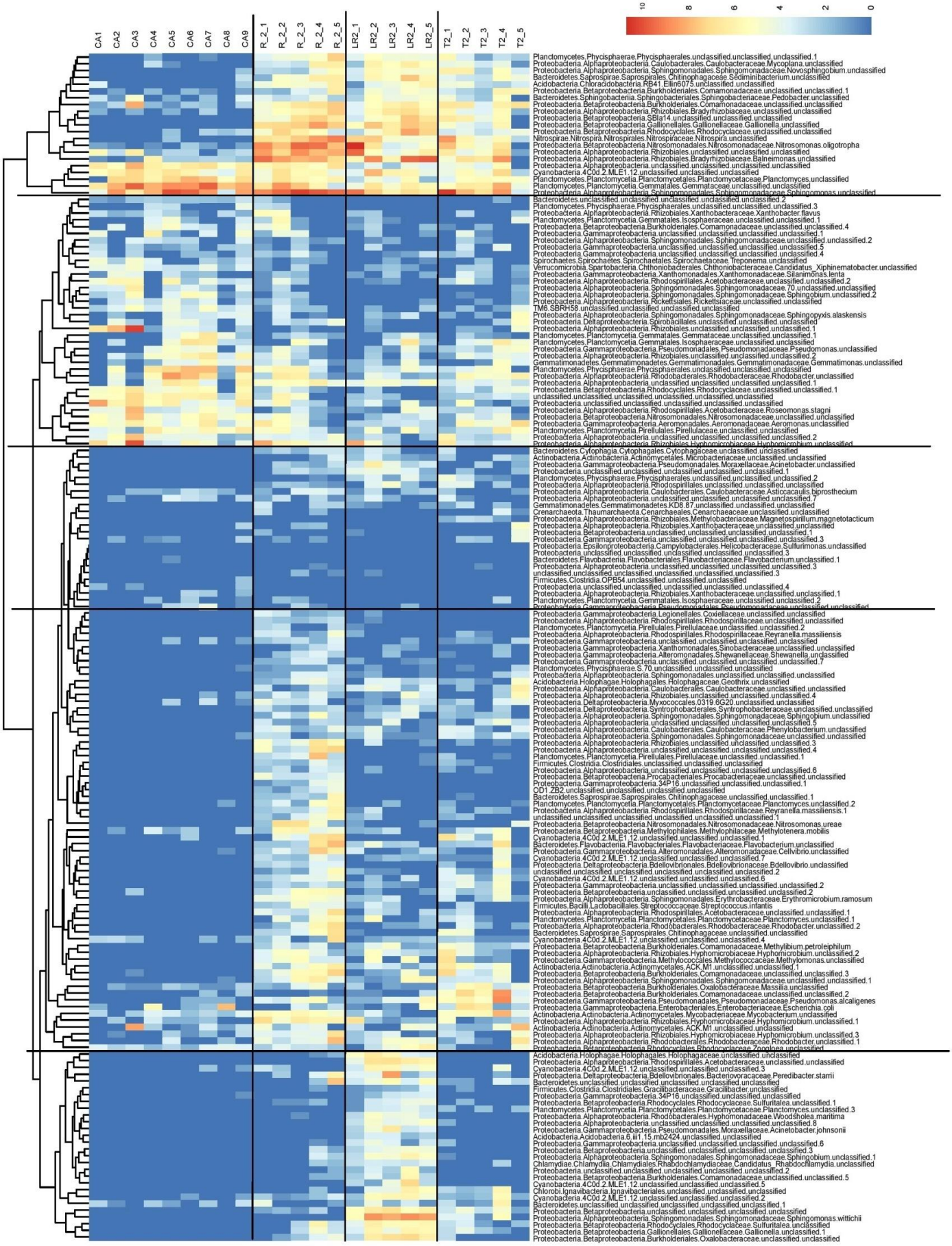


Figure 5.18 Heatmap showing the distribution of the most dominant OTUs in the system.

Community diversity

Statistical analysis showed insignificant temporal differences (Table C15-C18) between the different sampling points for the overall alpha diversity measures such as species observed (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J). Overall significant differences were however, observed between the sampling points for observed species (S_{obs}) and the Shannon diversity index (Table 5.9). When these differences between the points were further investigated by focusing on individual points (Table 5.10), significant differences in S_{obs} between R2 (main reservoir) and CA (water directly after chloramination) as well as between T2 (consumer tap) and CA, were observed. Only the two sampling points CA and T2 (furthest apart) showed a significant difference in terms of the Shannon diversity index.

Table 5.9 ANOVA table and P-values for observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J), evaluating an overall sample points statstical significant effect in community

Observed species (S_{obs})					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	3	1586.2	528.72	7.2913	0.001218
Within the groups	24	1740.3	72.51		
Total	27				
Shannon diversity index (H')					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	3	3.2221	1.07405	3.3787	0.03474
Within the groups	24	7.6293	0.31789		
Total	27				
Pielou's evenness (J)					
	DF	**SS**	**MS**	**F-value**	**P-value**
Between the groups	3	0.05882	0.019607	1.2105	0.3273
Within the groups	24	0.38875	0.016198		
Total	27				

Table 5.10 Tukey's post-hoc HSD test p-values for observed species (S_{obs}), Shannon diversity index (H') and Pielou's evenness (J), evaluating sampling points statstical significant effect in community

Sample comparison	**Observed species (Sobs)**	**Shannon diversity index (H')**	**Pielou's evenness (J)**
LR2-CA	0.15498	0.523724	0.84197
R2-CA	0.029615	0.186453	0.486738
T2-CA	0.001601	0.039373	0.40555
R2-LR2	0.906066	0.934935	0.955165
T2-LR2	0.349837	0.608598	0.920601
T2-R2	0.740652	0.913703	0.999377

In terms of the beta diversity metrics, significant spatial differences were observed for all the sample points, based on all four the metrics calculated using the actual OTU abundance data (Table 5.11). When the points were compared separately (Table 5.12), it again showed that there were significant spatial differences between all the individual sampling points. This was again supported by all four beta diversity metrics (Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac). The only exception was that the weighted UniFrac measure was not significantly different between LR2 (local reservoir) and T2 (tap). However, no significant seasonal (temporal) variations were observed (Table C19).

Table 5.11 AMOVA table for distance matrics on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all sampling points (CA-LR2-R2-T2), showing significant differences to one another (shown with a *)

Bray-Curtis					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.60785	3	0.869285	3.08822	<0.001*
Within	6.75561	24	0.281484		
Total	9.36347	27			
Jaccard					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.44816	3	0.816055	2.73473	<0.001
Within	7.1617	24	0.298404		
Total	9.60987	27			
Unweighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	2.41481	3	0.804935	3.29041	<0.001*
Within	5.87113	24	0.244631		
Total	8.28594	27			
Weighted UniFrac					
	SS	**DF**	**MS**	**Fs**	**P-value**
Among	1.94166	3	0.647218	4.77359	<0.001*
Within	3.254	24	0.135583		
Total	5.19565	27			

Table 5.12 AMOVA showing distance matrices on Bray-Curtis, weighted UniFrac, Jaccard and unweighted UniFrac on all sampling points in relation to one another, showing where the significant difference were found (shown with a *)

Sample comparisons	Bray-Curtis	Weighted UniFrac	Jaccard	Unweighted UniFrac
CA-LR2	<0.001*	<0.001*	<0.001*	<0.001*
CA-R_2	0.002*	<0.001*	<0.001*	<0.001*
CA-T2	0.018*	<0.001*	<0.001*	<0.001*
LR2-R_2	0.038*	0.001*	0.001*	<0.001*
LR2-T2	0.024*	0.086	0.01*	0.005*
R_2-T2	0.024*	0.038*	0.009*	<0.001*

The PCoA plot (Figure 5.19) showed that most of the samples grouped according to their specific sampling locations. A few outliers were also observed. As expected for such a short period of time a seasonal impact was not observed amongst the data, apart from amongst the CA samples where most of the spring samples clustered separately from those collected during winter.

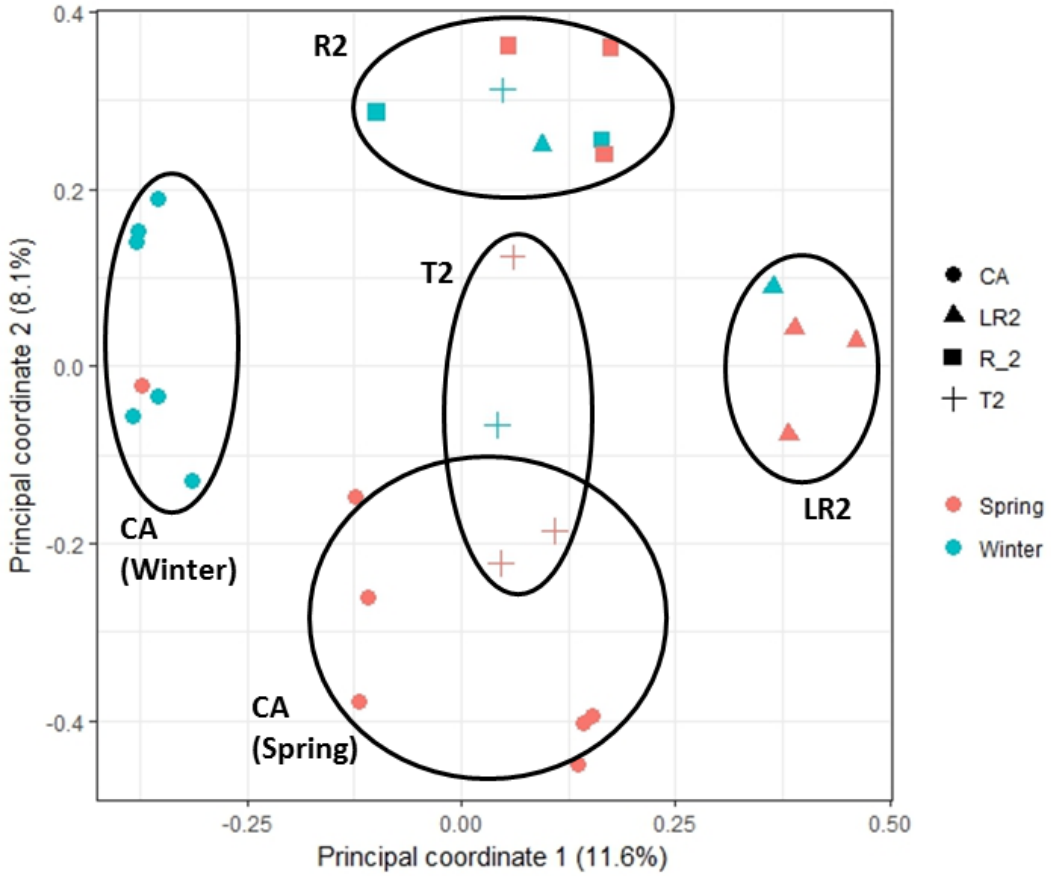


Figure 5.19 PCoA plot showing the community relation of all five sampling points over 6 months with corrosponding seasonal changes.

5.4 DISCUSSION

5.4.1 System 1 – An alternative treatment process

The dominance of *Proteobacteria* has been a common feature of drinking water studies and this system was no different (Pinto *et al.*, 2012; Lautenschlager *et al.*, 2014). In system 1, *Bacteroidetes* were observed to have the highest mean relative abundance for the raw water (A1). Again this was not unexpected as these bacteria are generally found in environments characterised by high levels of degradable organic matter. In addition, members of *Bacteroidetes* are known to attach themselves to particles and the surfaces of algae cells and the source water used for this drinking water system (system 1) contained high levels of algae. As the water was treated, the abundance of the OTUs linked to *Bacteroidetes* decreased substantially. *Cyanobacteria* were also seen at high abundances at A2, A4 and A5. These photosynthetic bacteria are commonly found in water environments and can adapt to most ecosystems (Scott, 2017). This treatment system showed traces of this group of bacteria throughout the system due to its initial presence in the source water.

Proteobacteria can be divided into 5 different classes. These classes are often associated with different environmental conditions. The *Alphaproteobacteria* is commonly dominant in drinking water systems, as these bacteria can grow at low levels of nutrients and can fix nitrogen as diazothrophs (Margulis *et al.*, 1988). *Betaproteobacteria* are highly diverse and are known to have large range of metabolic properties (Garrity *et al.*, 2003). *Gammaproteobacteria* is another class reported to be dominant in drinking water. It is the largest class in terms of species and high numbers are typically observed in anoxic environments. Clear waters with anoxic bottom layers provide perfect conditions for their growth (Stackebrandt*et al.*, 1988). *Deltaproteobacteria* are generally contributors to anaerobic sulphur cycles in water, and are found at lower abundances and lastly *Epsilonproteobacteria* are mostly found at very low abundances in drinking water systems.

Gammaproteobacteria had the highest mean relative abundance across all sample points in system 1. This is a common inhabitant of drinking water treatment plants (Lui *et al.*, 2016) but it is often more common to find the *Alpha*- and *Betaproteobacteria* to dominating the drinking water microbiome. As *Gammaproteobacteria* thrive in anoxic environments with enough light, both the original source water and the various stages within the treatment plant provided an optimal location for their growth. *Alphaproteobacteria* had the second highest mean relative abundance for most sample points (A1, A2, A3 and A5). As *Alphaproteobacteria* can grow at lower levels of nutrients it was not unexpected to observe them in this system or to notice an increase in their numbers during distribution (A4 to A5).

Significant spatial differences but insignificant temporal differences were reported for this system. The spatial differences were not unexpected as each of the sampling points represented a different section of the treatment process which would have a noticeable impact on the diversity of the community. The highest richness (S_{obs}), diversity (H') and evenness (J) were observed for the raw water (A1). Thereafter, these values typically decreased due to the impact of treatment. The significant differences were reported when the raw water was compared with the water after treatment and final chlorination (A1 vs A4 and A5). As the overall

characteristics of the community is strongly impacted by the associated treatment steps it was also not unexpected that temporal changes showed no significant impact.

Beta diversity was examined to evaluate the diversity differences between communities based on the specific OTU occurrence and abundance data. As anticipated, significant spatial differences were reported between communities at different sampling sites when measured during the same month. Therefore, the different treatment steps not only determine how the community diversity will be structured (alpha diversity) but it also determine which bacterial species (OTUs) will survive and proliferate (OTU abundance) under the specific conditions created in the system. The only exception to these reported changes in the community membership and structure was notice for sampling points A4 (after treatment and final chlorination) and A5 (distribution of the treated water). As no further treatment was applied between these two points, no significant differences were reported for their beta diversity distances. This observation was also supported by the Principal Coordinates Analysis (PCoA). The insignificant changes during distribution was also noticed in other studies (included in Chapter 4), which showed that the bacterial community within the distribution systems could remain fairly constant in terms of membership and structure especially when associated with a relatively short distribution line (65 km) and no reservoirs within this part of the network.

Significant temporal changes could be observed between some of the communities associated with the same sampling point but sampled at a different time point. When samples were grouped according to seasons, some significant differences, in the beta diversity metrics, were reported. The main changes were observed between autumn and spring, where there was an estimated 6-month difference between the sampling events. A similar pattern was observed when samples were compared based on a month to month basis. Both structural and membership metrics showed insignificant changes on a month to month basis unless these months were at least three or more months apart from one another.

The flow cytometry results indicated that the treatment plant was working efficiently as there was a decrease in cell counts from the raw water to the water in the distribution system, with 4 597 cells/ml in the raw water to 319 cells/ml in the distribution system. This is to be expected as the treatment plant is meant to control the number of bacteria in the distribution system. Reporting of actual cell counts are important, as there is a need to distinguish between intact and damaged cells (Alan *et al.*, 1988). The high number of damaged cells in the raw water is common as the surface water is highly contaminated and several environmental factors such as sunlight and competition could lead to cell death. The data also demonstrated that pre-treatment was effective in removing most of these cells.

5.4.2 System 2 – Groundwater system

In the groundwater system as many as 68 different phyla were observed across all sampling points over the 11 month period but only a small number bacteria dominated. This was clearly demonstrated when the distribution of the dominant OTUs (abundance of > 1%) was plotted using a heatmap. As for most other drinking water systems the *Proteobacteria* again dominated the system with a high mean relative abundance

of between 42-59%. The *Gammaproteobacteria* again dominated at most of the sample points (F1, F2 and F4), which is not typical for all systems studied. In this system, which received limited treatment, the overall composition of the communities were fairly constant apart from the community associated with the reservoir (F3). At this point most of the *Proteobacteria* were represented by the *Alphaproteobacteria* and specifically by unclassified OTUs linked to the order *Rhizobiales* and these bacteria may be involved in nitrogen cycling. The reservoir also had a high abundance of members of *Firmucutes* (*Actinobacteria*). These OTUs included genera such as *Bacillus* and *Mycobacterium*, all known to survive under extreme conditions such as those created by high oxidative stress caused by disinfection.

No significant differences between communities associated with different sampling points were reported based on the all three of the alpha diversity metrics. As mentioned above, this system received limited treatment and based on the lack of factors that could shape the overall characteristics and structure of the community this was not unexpected. In addition, no significant temporal changes linked to seasons of months were also observed.

The analysis of the beta diversity metrics also showed no significant differences that could be linked to seasonal changes. As water came from a well-protected aquifer, seasonal variation in the quality of the raw water would be unexpected. As 57% of the daily flow originates from F1, it is also expected that this source would have a greater impact on the overall community, but no significant spatial differences were observed between most of the sampling points in the system. As this is a low biomass system, confirmed by the flow cytometry data discussed below, it is expected that slight changes in bacterial levels could impact on the structure or membership of the communities associated with the system. Based on the Principal Coordinates Analysis, most of the samples (except samples from the reservoir (F3)) clustered together. It is uncertain why several other parameters also indicated that the reservoir community was unique compared to the other sampling sites. It should be investigated whether the samples collected from the reservoir were truly representative of the bulk water in the reservoir.

Flow cytometry tests revealed low total cell for both of the groundwater sources, F1 (11 cells/ml) and F2 (12 cells/ml). The bacteria levels were slightly higher at the reservoir F3 (189 cells/ml) but had similar levels to the raw water when measured at the tap of the consumer. This indicated that the groundwater sources and water that was distributed, was of excellent microbial quality as is often the case with groundwater systems (Lui *et al.*, 2013).

5.4.3 System 3 – Distribution and reticulation system

Proteobacteria again had the highest mean relative abundance across all sample points within the distribution and reticulation system. As seen in other studies (also Chapter 4), the community present in the water directly after chloramination differs from the community present in the downstream distribution system. During this part of the study an increase in the relative abundance of the *Proteobacteria* was observed after chloramination.

This coincided with an increase in the relative abundance of the *Betaproteobacteria,* which is not uncommon for many distribution systems.

Significant structural differences (based on alpha diversity) were observed between communities samples directly after chloramination, compared to those associated with the main reservoir or the tap of the consumer. Apart from the above-mentioned changes based on community diversity characteristics, significant spatial changes were also observed between all the sampling points as was reflected by the structure and membership data for all the communities compared. The PCoA plot also indicated that the communities linked to specific sampling points were unique. Due to the short sampling period no temporal changes were observed based on any of the beta diversity metrics. In order to observe such changes sampling should be conducted over a longer time period. The PCoA plot indicated same seasonal differences between samples collected during chloramination (CA).

5.5 SUMMARY

The primary aim of this phase of the study was to evaluate whether the molecular approach utilised during the main project (Chapter 3 and 4), would also be applicable to and benefit South African communities supplied by smaller drinking water and reticulation systems. For this reason, systems were included in this study, which used different treatment technologies (pre-chlorination and a DAF/Filtration system) and/or water sources (groundwater). From the data presented in this chapter it is clear that the approach to study the microbial ecology of water treatment and distribution systems using 16S profiling is also applicable to alternative systems. This approach provided valuable information on the impact of treatment, distribution, seasons and other water quality parameters on the final microbial quality of the water supplied to the consumer, especially when alternative treatment approaches or water sources are used.

This study clearly showed that the levels of some bacteria may be enhanced by different treatment steps. The exact processes involved is unclear, as many of the OTUs with increased relative abundance levels due to treatment, represent unclassified bacteria. To investigate these interactions in greater detail, a metagenomic approach should be applied to recover the genomes of these bacteria and potential functional and metabolic capabilities. Although the level of certain OTUs were enhanced during treatment, the disinfection step remained the most important driver for the observed shifts in the microbial community present in the final water. During this step most of the bacteria, that were associated with specific treatment steps, were again removed from the system.

This study again confirmed that groundwater, from a well-protected aquifer, remains a source of drinking water as microbes are present at very low levels. It was shown that for such systems only minimal disinfection and treatment are required to ensure safe drinking water. Attention should however, be given to the distribution of this water as no residual disinfectants are present to counter contamination events.

A further aim of the current study was to determine the impact of the drinking water reticulation system, characterised by smaller diameter pipes, on the bacterial community in the water. The study clearly showed that at each step within the distribution and reticulation system had an impact on the quality of the water that reaches the consumer. Further studies to understand these systems are recommended. This can be done including several final distribution points (taps) during the evaluation and ensuring that the sampling should be done over a period of at least one year.

CHAPTER 6: DIVERSITY AND IDENTITY OF POTENTIAL BACTERIAL PATHOGENS IN A LARGE-SCALE DRINKING WATER DISTRIBUTION SYSTEM

6.1 INTRODUCTION

Safe drinking water is considered a basic human right and is essential to human health. To protect the consumer, drinking water is often subjected to a range of treatment processes to improve the microbial quality and safety of the water. The safety of drinking water is however, not only determined by the quality of source water and the subsequent treatments thereof but also by distribution practises and conditions (Berger *et al*., 2015). The focus of managing water quality in a drinking water distribution system is to control the growth of microorganisms as well the occurrence of pathogenic microorganisms that could pose a potential health risk (WHO, 2004 and 2011). The pathogens present in drinking water distribution systems are either waterborne or water-based environmental pathogens.

Enteric waterborne microorganisms come into contact with water sources primarily due to contamination with human or animal faecal matter (WHO, 2011), while some also have the ability to form part of the natural community present in fresh water systems (Newton *et al*., 2011). The water-borne (enteric) pathogens have the ability to cause several diseases including gastroenteritis. The severity of the infection is determined by the virulence of the pathogen but it can also be influenced by environmental conditions, the immune system of the host as well as contributing factors as a result of water treatment processes, i.e. chlorination and chloramination (Brettar and Hofle, 2008).

Water-based environmental pathogens typically thrive in the aquatic ecosystem and drinking water is suggested to be a vehicle for the transmission of these pathogens as they often survive treatment and can regrow in the distribution system (Van der Wielen and Van der Kooij, 2013). These environmental pathogens are suggested to be opportunistic based on their ability to only infect susceptible individuals. These microorganisms have weak virulence but have retained the ability to infect immunocompromised humans and initiate disease (Van der Wielen and Van der Kooij, 2013). Individuals infected are typically the old and young members of society who are incapable of fighting disease. Other individuals could also be susceptible due to prior infections or conditions (HIV, cystic fibrosis, cancer, etc.) that have led to immunodeficiency. These pathogens come into contact with humans via ingestion, direct contact with drinking water or through inhalation of aerosolised pathogens. Some of these pathogens are also spread through other sources such as food and soil.

The detection of bacterial pathogens in water could either be based on culture-dependent or culture-independent (molecular) approaches. Culture-dependent methods include heterotrophic plate counts,

membrane filtration (MF) and multiple tube fermentation (MTF) tests amongst others (Douterelo *et al*., 2014). At present the use of culture-independent molecular methods include different types of hybridisation or polymerase chain reaction (PCR) targeting specific pathogens or high throughput sequencing to study the whole community (Wang *et al*., 2017). The analysis of the overall microbial community linked to phylogenetic analyses provides an ideal opportunity to obtain information on the presence and abundance of potential bacterial pathogens within the drinking water distribution system.

In this study, the focus was on the bacterial community in a drinking water distribution system as determined by 16S rRNA community profiling. 16S rRNA profiling involves the PCR amplification of a variable region of the 16S gene using the total genomic DNA extracted from the sample as a template. The amplified mixture of 16S amplicons is then sequenced with high throughput sequencing (Illumina MiSeq) (Vaz-Moreira *et al*., 2012 and 2013). This approach provides a broader basis and overview of the community within a drinking water environment and allows the researcher to make deductions on the trends observed. The sequences obtained can also be used for further analyses against type strain sequences to improve the identification of these sequences.

The purpose of this study was to determine the presence and abundance of potential bacterial pathogens in a large water distribution system in South Africa by performing 16S community profiling. Sequence data of the bacterial community in the system was used to identify operational taxonomic units (OTUs) that are closely related to or match the sequences of the type strains of potential water-borne and opportunistic water-based environmental pathogens. The study will provide an understanding of the potential bacterial pathogen community found in the distribution system studied. The study will also provide insight with regard to the health risk posed by the potential pathogens identified.

6.2 MATERIALS AND METHODS

6.2.1 Sampling and sample processing

Samples and resulting 16S rRNA community profile data analysed in this study were obtained from the data presented in Chapter 4. The sampling points were: Cl (chlorinated water directly after chlorination), BfAm (chlorinated water after distribution), ClAm (water directly after chloramination), DSP1; DSP2; DSP3; DSP4 and DSP5 (distribution system) (Figure 6.1).

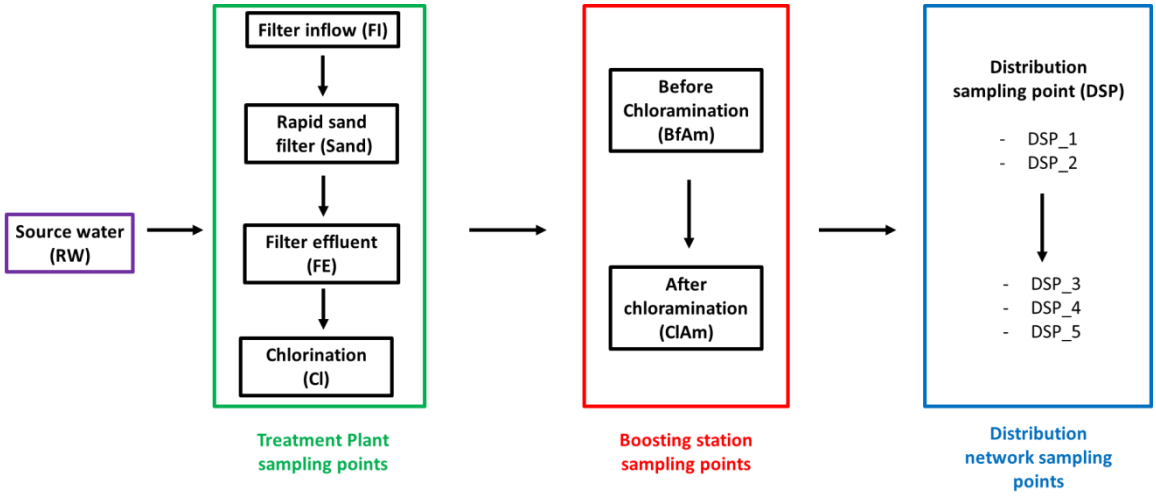


Figure 6.1 A schematic layout of the treatment works and DWDS points sampled during this study.

6.2.2 Illumina sequencing and sequence processing

The V4 hypervariable region of the 16S rRNA gene was targeted following the method involving dual-index paired-end sequencing of the region which produces 250 bp fragment lengths as described by Kozich and colleagues (2013). The raw sequence data were quality trimmed and checked for chimeras using MOTHUR v.1.35.1 (Schloss *et al*., 2009). The remaining quality filtered sequences were classified using the Greengenes database (DeSantis *et al*., 2006), with a threshold confidence level of 80%. Sequences assigned to an unknown taxonomy domain level were discarded. Sequences were further assigned to OTUs at a 97% similarity.

Representative sequences of the OTUs corresponding to families and genera known to be potential bacterial pathogens in water (Jordaan, 2015) were extracted from the complete list of OTU sequences retrieved from MOTHUR. The OTU sequences were classified according to the phyla, order, family and genera. OTUs associated with families or genera previously detected as pathogens in drinking water distribution systems were selected from this database. These selected OTUs were separated into two tables, namely all OTUs classified up to the family level and all OTUs classified up to the genus level.

6.2.3 Taxonomic Sequence analysis

Phylogenetic analyses

Sequence files containing the representative OTU sequences for each family and genus were constructed using BioEdit Version 7.2.5. Reference sequences of both obligate and potential pathogens, which have been detected in drinking water and/or distribution systems, were obtained from the contaminants reference list constructed in a previous study from the EPA's Candidate Contaminants List 3, World Health Organization (WHO) microbial fact sheet (2011), as well as previous literature (Jordaan, 2015). The OTU sequences of

each family were combined with the reference sequences of the pathogenic bacterial strains. The alignment of the sequence files were performed using MAFFT Version 7 online (http://mafft.cbrc.jp/alignment/software/).

Maximum likelihood phylogenetic trees were constructed using MEGA Version 7.0.26 using 100 bootstrap replicates and edited in Inkscape 0.92. Upon evaluation, the OTU sequences shown to be closely related to the pathogenic bacterial type strains were extracted and separate files were constructed containing these OTUs and the relevant pathogenic type strains. A nucleotide BLAST using the NCBI database was also performed on the OTU sequences shown to be closely related to the pathogenic bacterial strains in order to obtain a putative identity for the OTU. The BLAST result consensus was used as a guide to obtain non-pathogenic bacterial type strains closely related to the pathogenic bacterial strains. Sequence files were constructed containing the OTUs closely related to the pathogenic bacterial strains together with the pathogenic bacterial strains as well as the non-pathogenic bacterial strains which were shown to be closely related to the pathogenic bacterial strains in the database. The previous studies, which contained phylogenetic trees showing the relatedness of the pathogenic bacterial strains to non-pathogenic bacterial strains were obtained from the list of Prokaryotic Names with Standing in Nomenclature (LPSN) database and the respective sequences were retrieved based on their accession numbers in NCBI Genbank as well as from the “All-species Living Tree Project” (LTP) database. The sequence files containing the OTUs closely related to pathogenic bacterial types, non-pathogenic bacterial type strains related to the pathogen and a relevant outgroup from the same or neighbouring genera was aligned using MAFFT Version 7 online (http://mafft.cbrc.jp/alignment/software/). Maximum likelihood phylogenetic trees with 100 bootstrap replicates of each family tree containing the OTUs and relevant pathogenic strains with outgroups from a different family were constructed using MEGA Version 7.0.26 and edited in Inkscape 0.92. These trees were the sub-trees of the combined trees previously constructed to analyse if the relevant OTU was pathogenic or non-pathogenic. The OTUs found to be pathogenic were analysed further based on their occurrence in the drinking water distribution system.

6.2.4 Relative abundance

The OTUs found to be potential bacterial pathogens upon analyses of the phylogenetic sub-trees were analysed further based on their abundance and occurrence at specific sampling points in the system. Relative abundance graphs for each of these OTU were constructed in Microsoft Excel 2013. The graphs represented the relative abundance of the respective OTUs across the eight sampling points in the twelve months sampled.

6.3 RESULTS

6.3.1 Taxonomic position of OTUs

The original number of OTUs present in the dataset was 8891. From this large dataset, OTUs were eliminated based on their classification. On the phyla level, 8110 OTUs were classified and 781 OTUs were eliminated as they could not be assigned to a known phylum. A total of 6113 OTUs were classified to an order level and 2778 OTUs were eliminated. Furthermore, classification of the OTUs to the family level left 4112 OTUs classified and 4779 OTUs were eliminated. From the 4112 classified OTUs, a subset of the database was created where the OTUs were classified into the families which contained potentially pathogenic bacteria using the contaminants reference list containing potential pathogens detected in DWTP and/or DWDS, constructed in a previous study (Jordaan, 2015). From the subset database, 677 OTUs associated with potential pathogens classified up to family level were identified (Table 1). Furthermore, amongst the 677 OTUs classified up to family level, 249 OTUs were further classified up to genus level (Table 2). Phylogenetic trees were then constructed using the OTUs from Table 1 and 2, respectively (677 OTUs).

6.3.2 Phylogenetic analyses

A total of 35 phylogenetic trees representing the families linked to OTUs listed in Table C1 and C2 (677 OTUs), were constructed. The families containing only three sequences (i.e. the pathogenic bacterial stain, the OTU and an outgroup) were excluded. These families included *Brevibacteriaceae, Francisellaceae*, *Pasteurellaceae* and *Promicrospiraceae*. Based on the phylogenetic trees constructed using the OTUs in Table C1 and C2, 40 OTUs, shown to group closely with the sequence representing the pathogenic bacterial strain, were further used to construct sub-trees. A total of 26 phylogenetic sub-trees were constructed containing the families listed in Table C3. From the sub-trees constructed, 18 OTUs were found to represent potential pathogens (from an original number of 8891 OTUs). The phylogenetic trees and sub-trees containing the potential pathogens are represented by Figures 6.2 to 6.16. The additional figures of phylogenetic trees showing OTUs that were not related to pathogenic strains as well as the additional phylogenetic trees consisting of OTUs which were not closely related to the pathogenic strain, were not included.

Four OTUs, namely (a) OTU136, (b) OTU168, (c) OTU1284 and (d) OTU2588 were closely related to waterborne pathogens. The pathogens which these OTUs may potentially represent included (a) *Escherichia coli* / *Escherichia fergusonii* / *Shigella sonnei* / *Shigella flexneri*, (b) *Streptococcus parasanguinis*, (c) *Yersinia enterocolitica* subsp. *enterocolitica* / *Serratia liquefaciens* / *Serratia plymuthica* and (d) *Enterococcus durans* / *Enterococcus faecalis* / *Enterococcus faecium* / *Enterococcus hirae*, respectively (some OTUs are associated with more than one closely related pathogen).

The fourteen OTUs closely related to water-based environmental pathogens are (a) OTU39; (b) OTU94: (c) OTU108, (d) OTU150, (e) OTU250, (f) OTU282, (g) OTU479, (h) OTU517, (i) OTU747, (j) OTU754, (k) OTU1172, (l) OTU1406, (m) OTU1654 and (n) OTU1997. The pathogens that these OTUs may potentially

represent belong to the families (a) *Sphingomonadaceae*, (b) *Aeromonadaceae*, (c) *Staphylococcaceae*, (d) and (g) *Methylobacteriaceae,* (e) *Enterobacteriaceae,* (f) *Moraxellaceae*, (h) *Corynebacteriaceae,* (i) *Micrococcaceae*, (j) *Bacteroidaceae*, (k) *Burkholderiaceae*, (l) *Mycobacteriaceae*, (m) *Bacillaceae* and (n) *Alcaligenaceae*. Phylogenetic trees of these families showing the placement of the specific OTU are shown below.

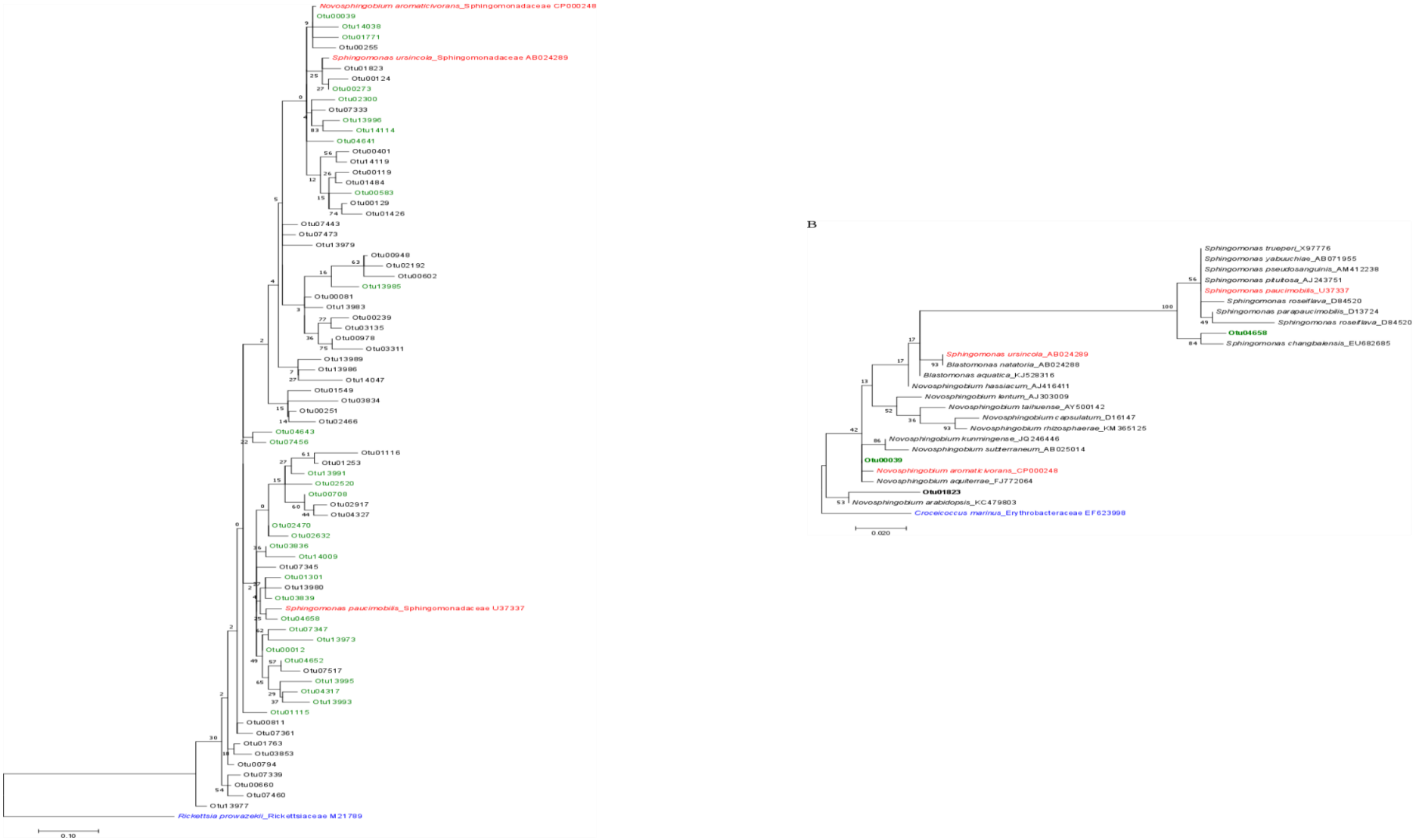


Figure 6.2 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Sphingomonadaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Sphingomonadaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

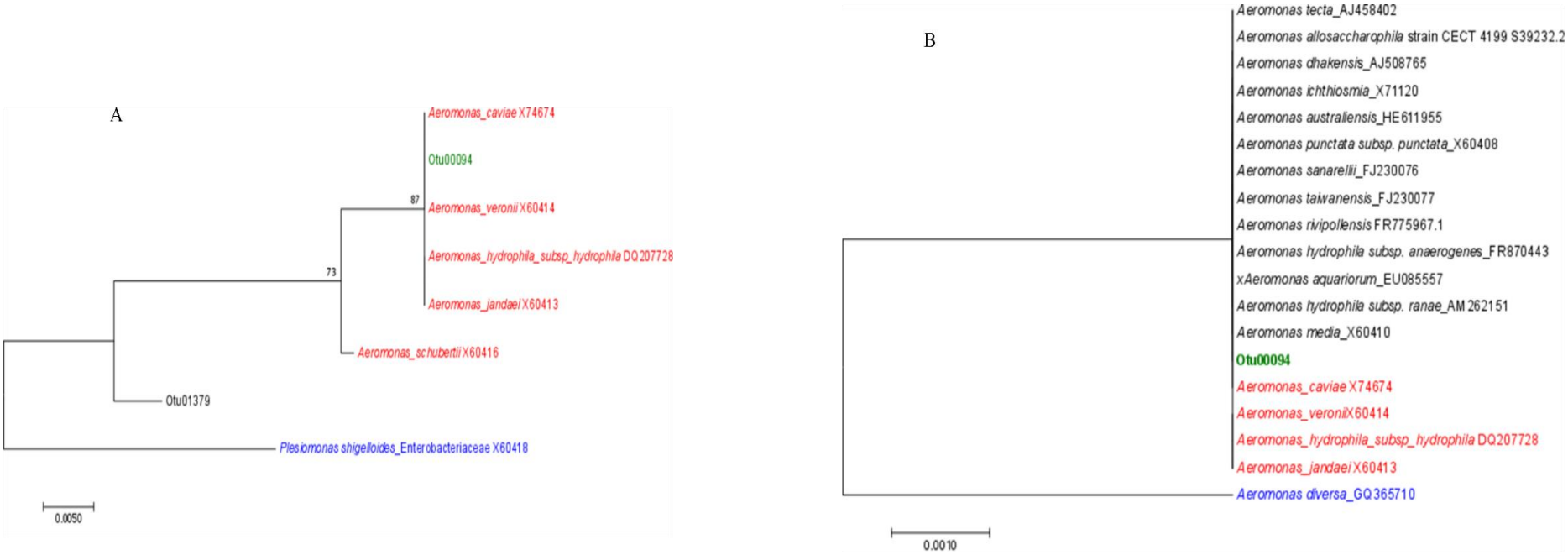


Figure 6.3 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Aeromonadaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Aeromonadaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

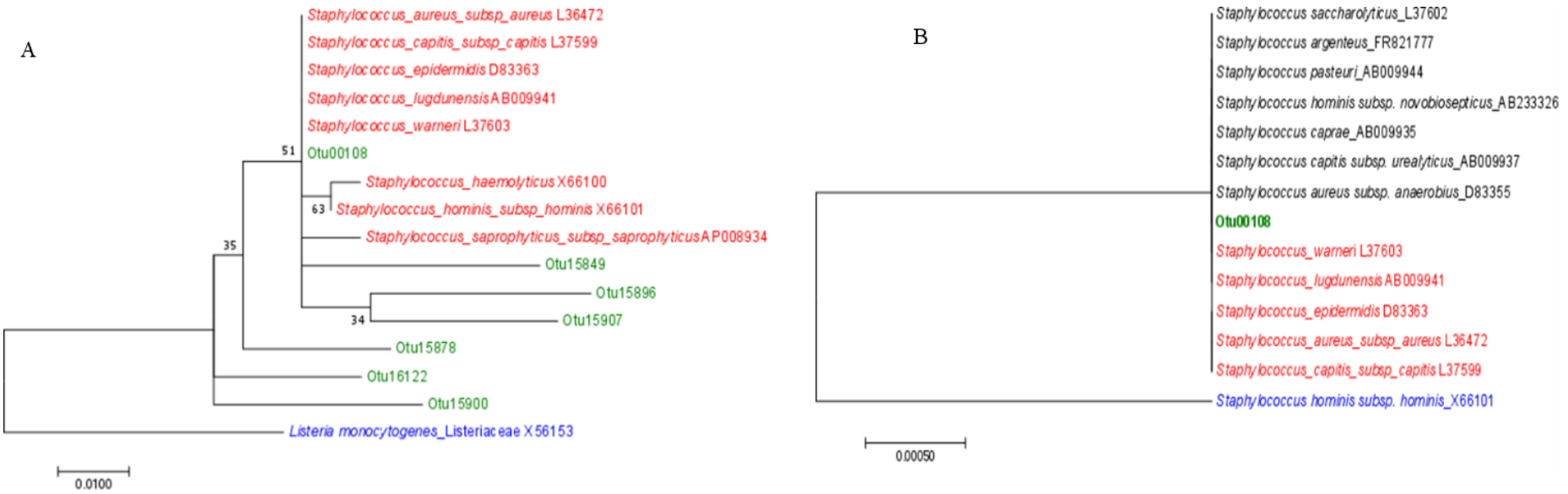


Figure 6.4 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Staphylococcaceae*. (B) Maximum likelihood tree representing the potentially pathogenic. OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Staphylococcaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

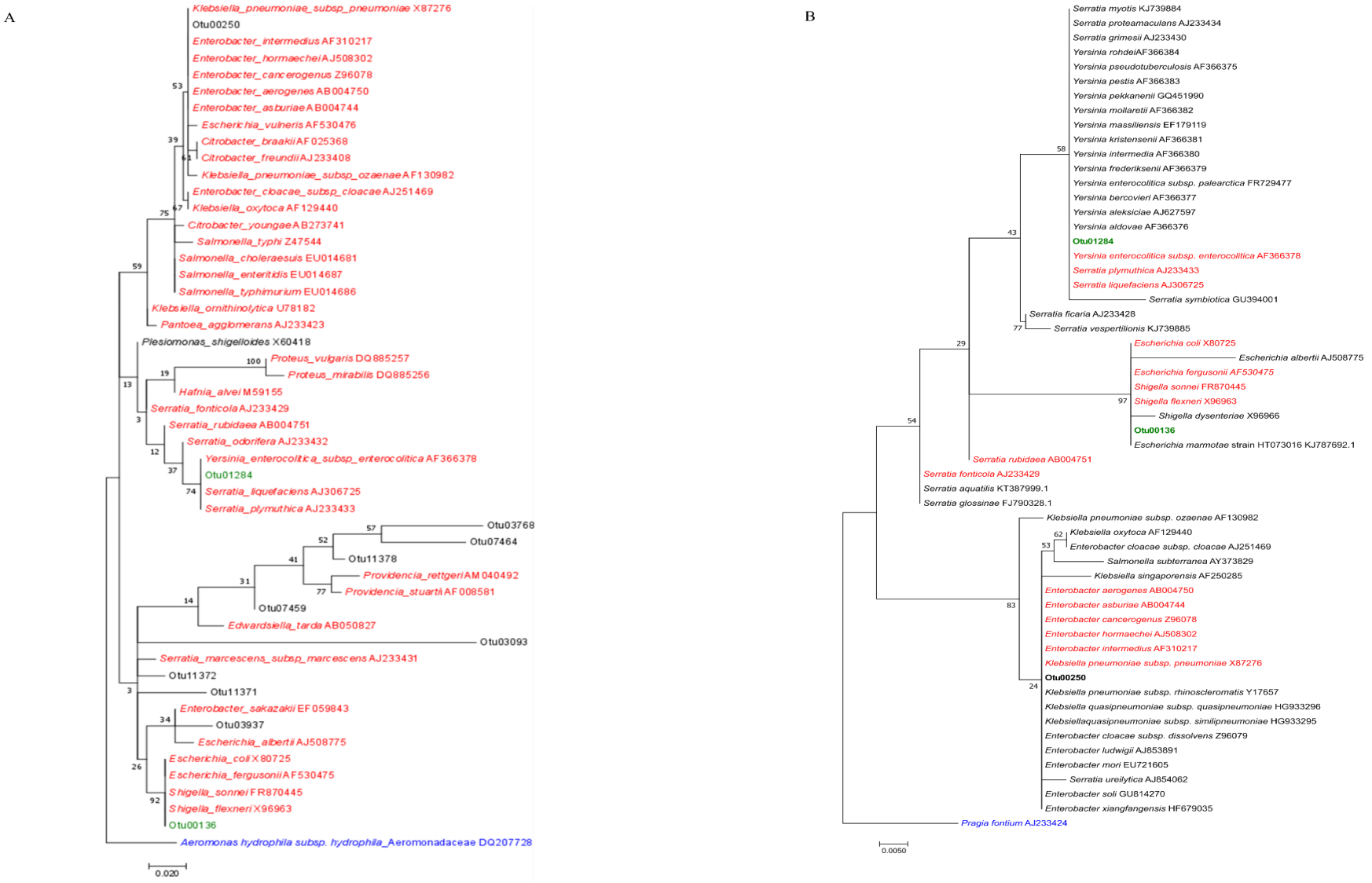


Figure 6.5 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Enterobacteriaceae*. (B) Maximum likelihood tree representing the potentially pathogenic. OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Enterobacteriaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

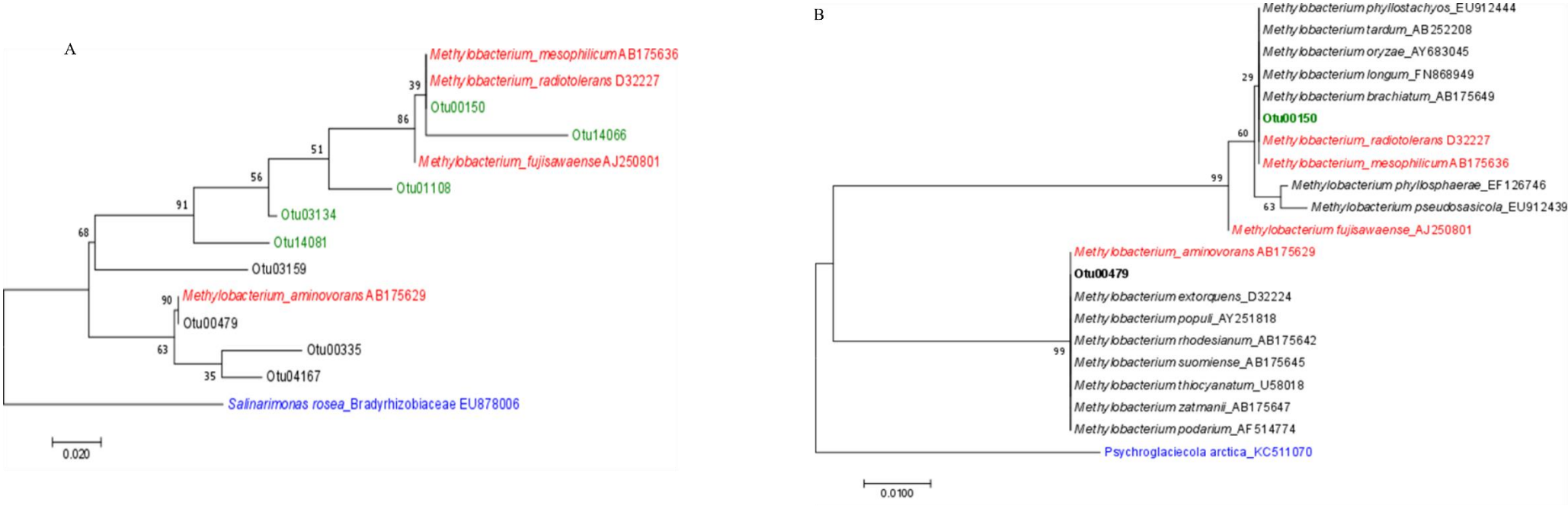
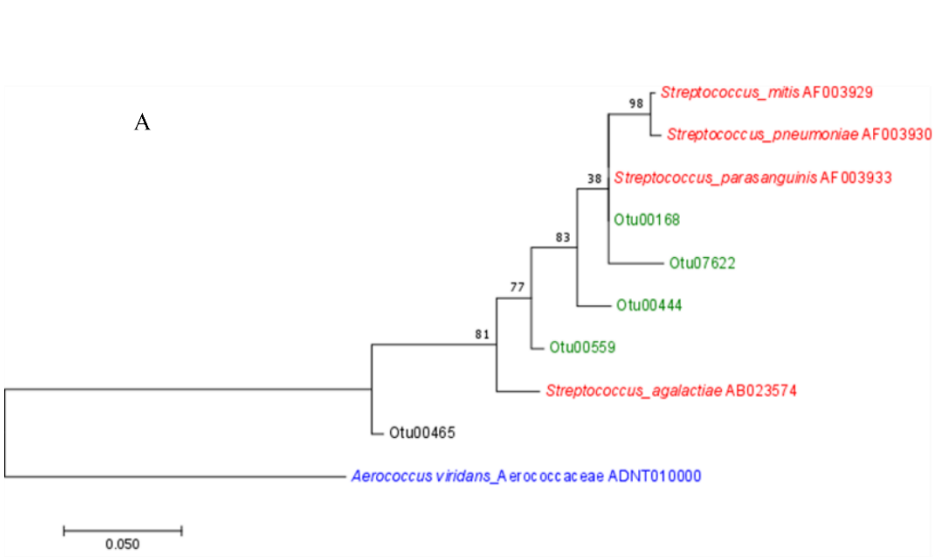


Figure 6.6 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Methylobacteriaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Methylobacteriaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.



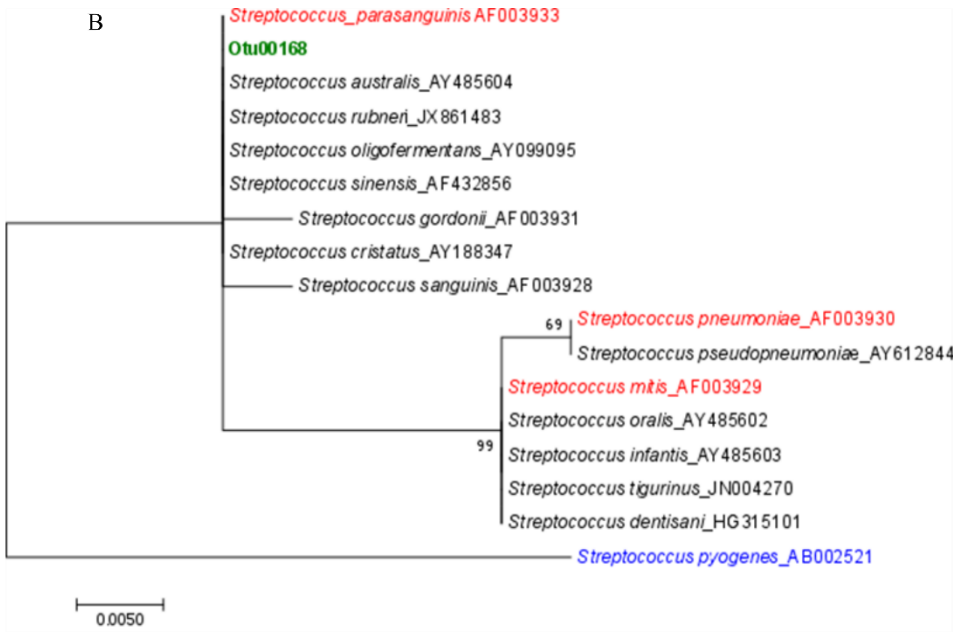


Figure 6.7 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Streptococcaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Streptococcaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

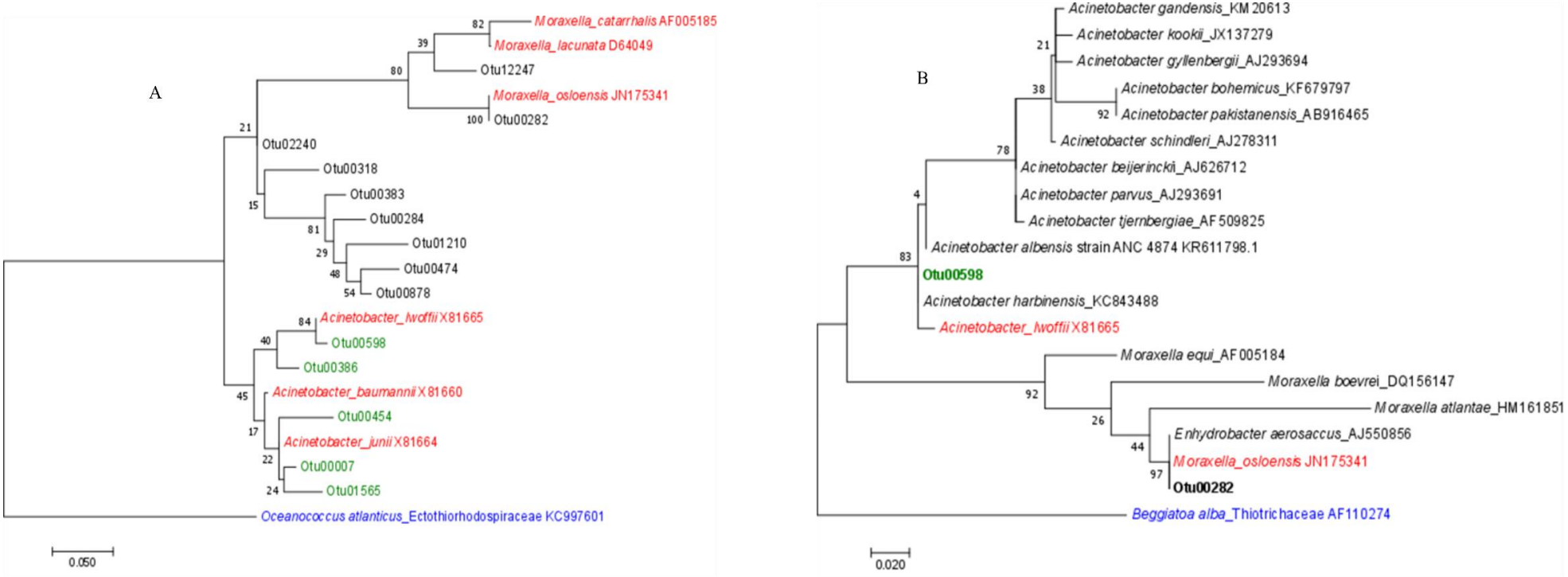


Figure 6.8 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Moraxellaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Moraxellaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

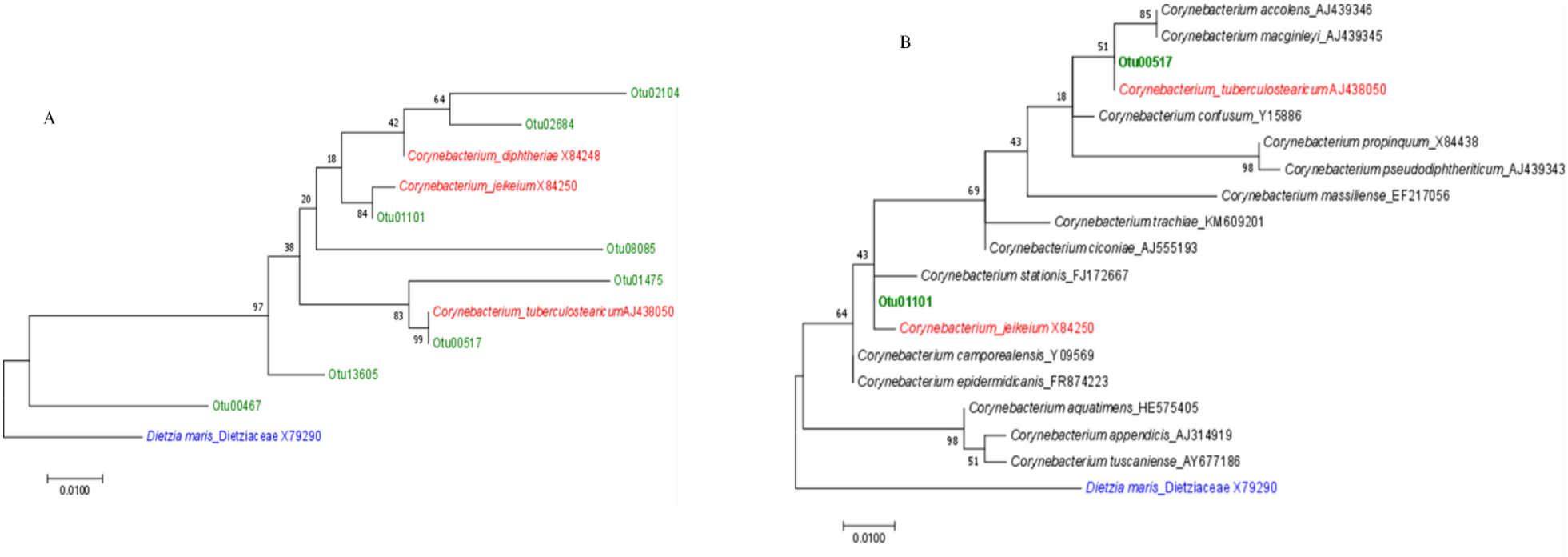


Figure 6.9 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Corynebacteriaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Corynebacteriaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

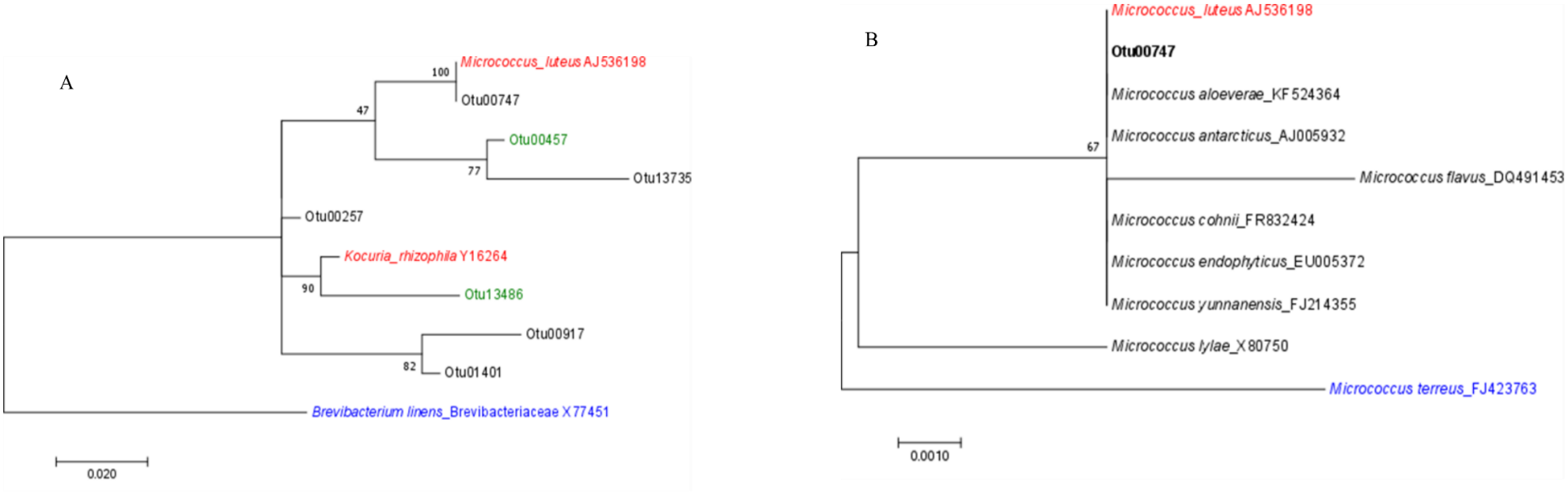


Figure 6.10 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Micrococcaceae*. (B) Maximum likelihood tree representing the potentially OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Micrococcaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

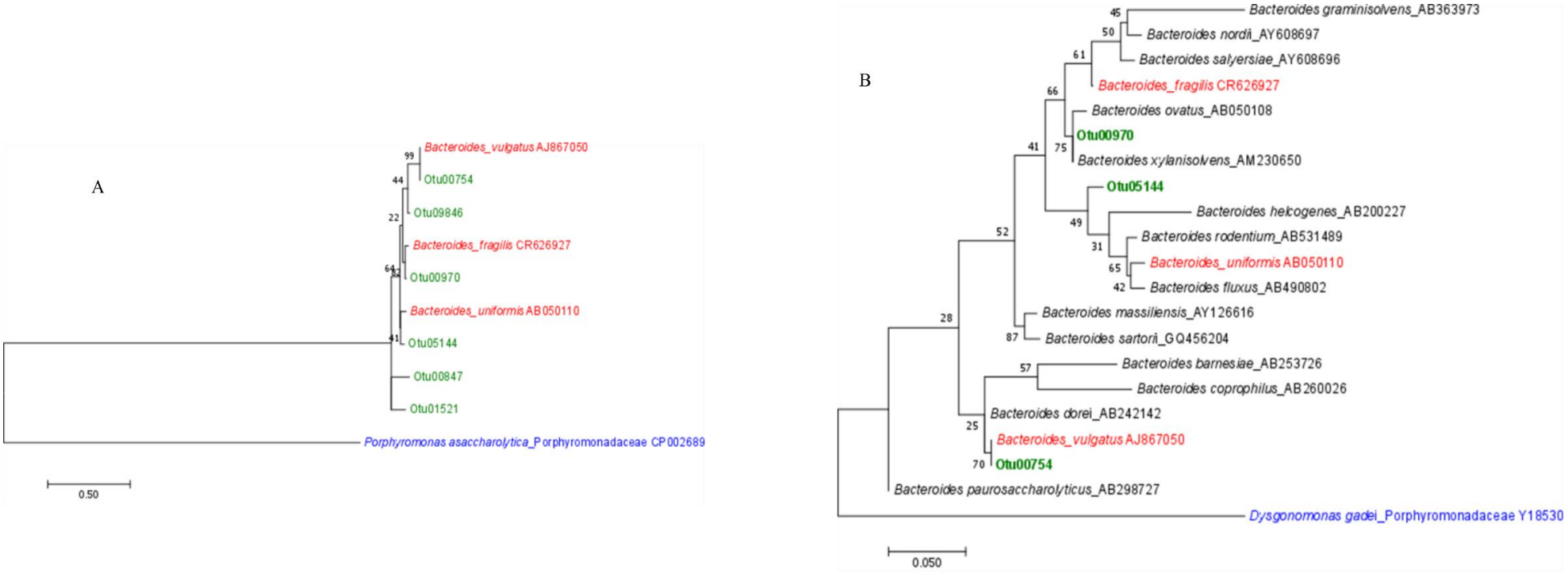


Figure 6.11 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Bacteroidaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Bacteroidaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

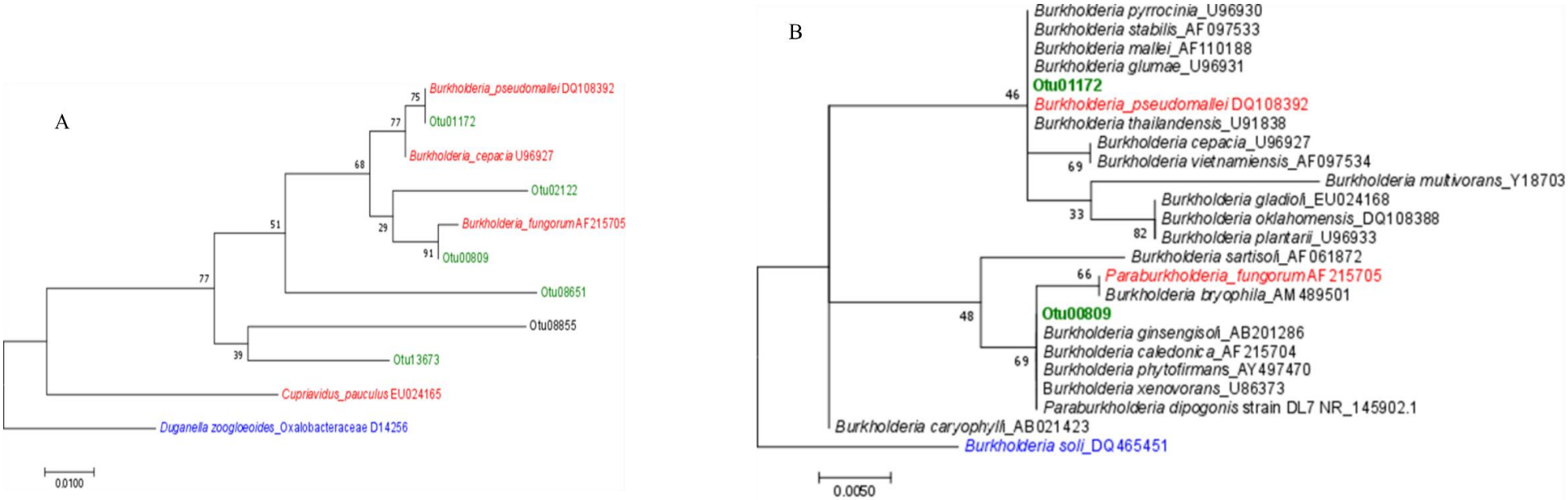


Figure 6.12 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Burkholderiaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Burkholderiaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

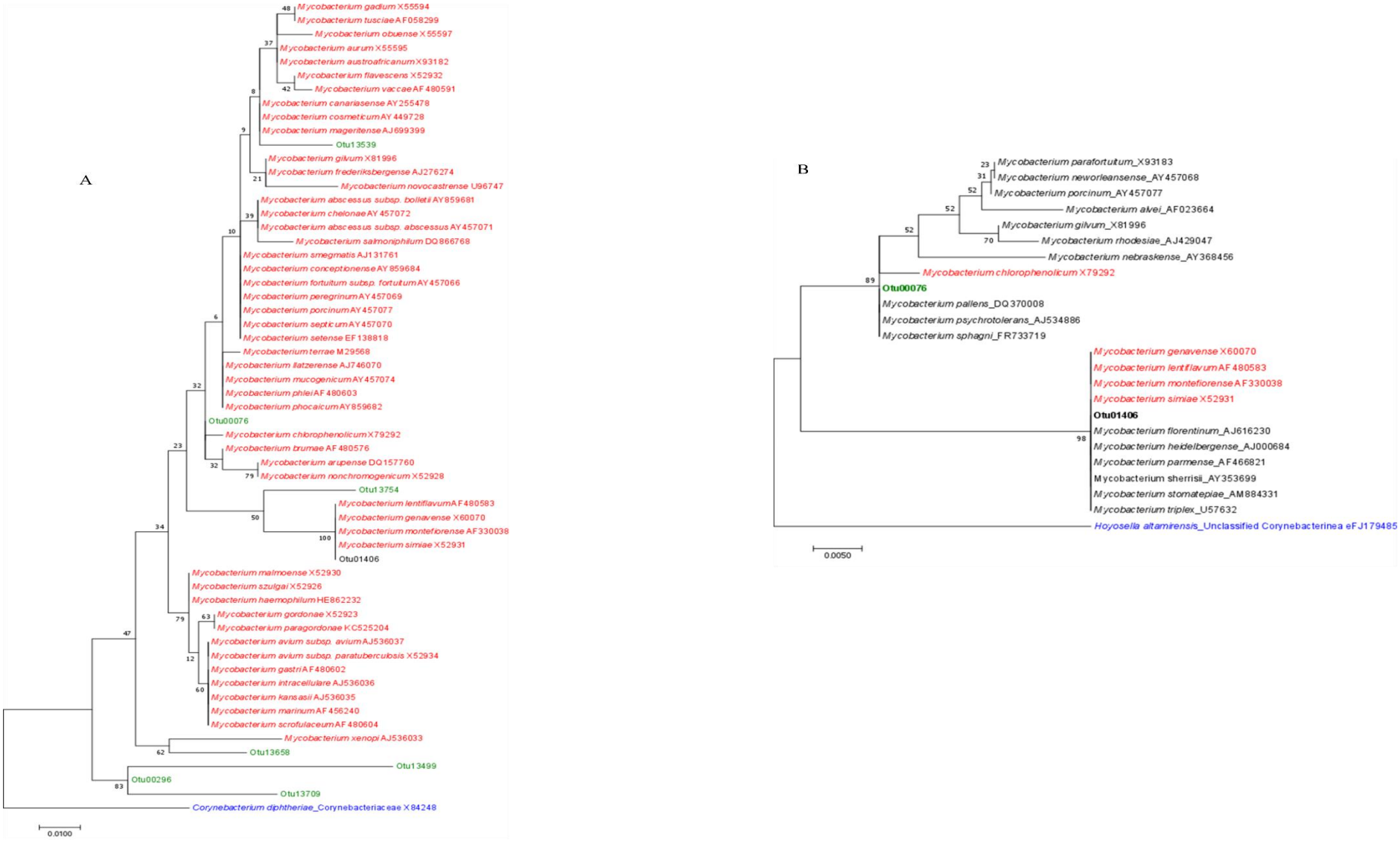


Figure 6.13 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Mycobacteriaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Mycobacteriaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

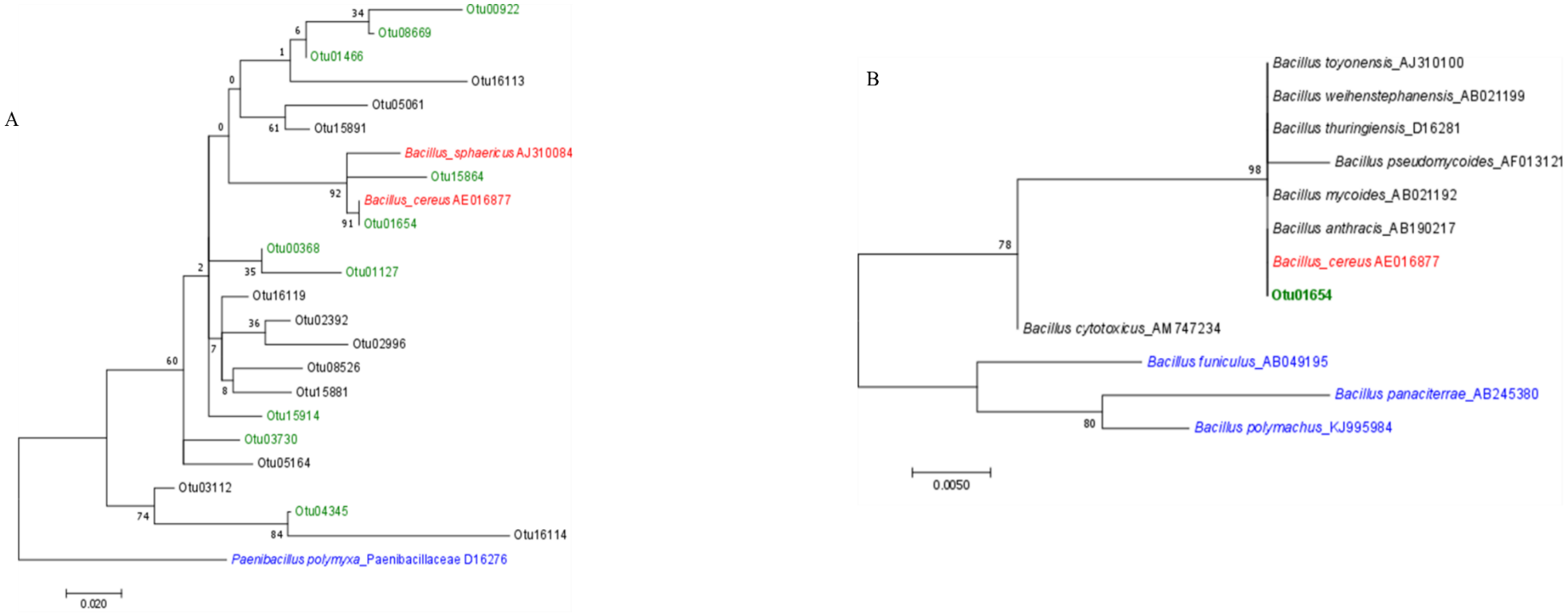


Figure 6.14 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Bacillaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Bacillaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

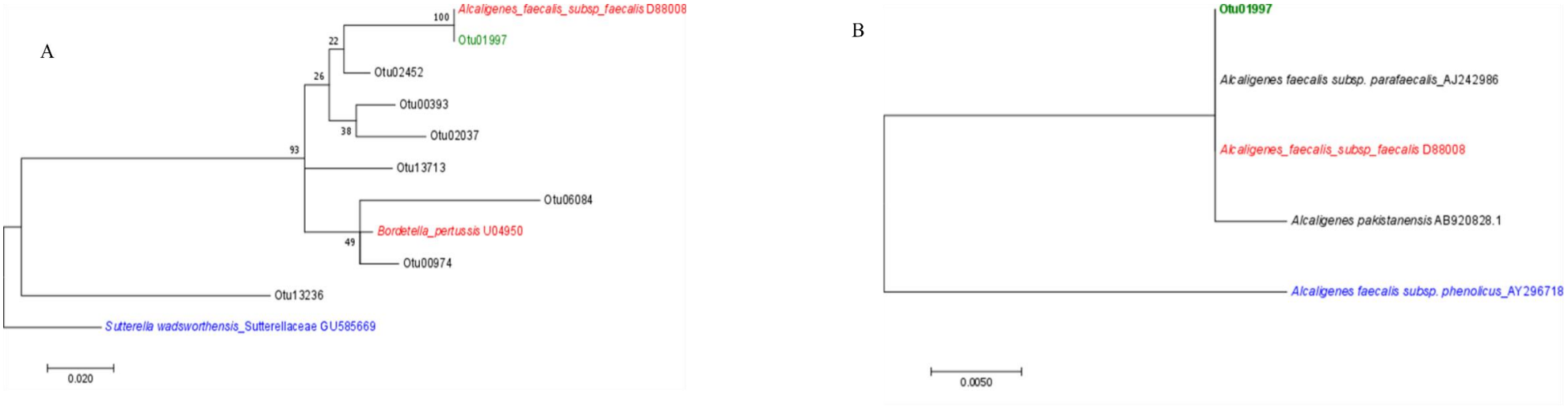


Figure 6.15 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Alcaligenaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Alcaligenaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

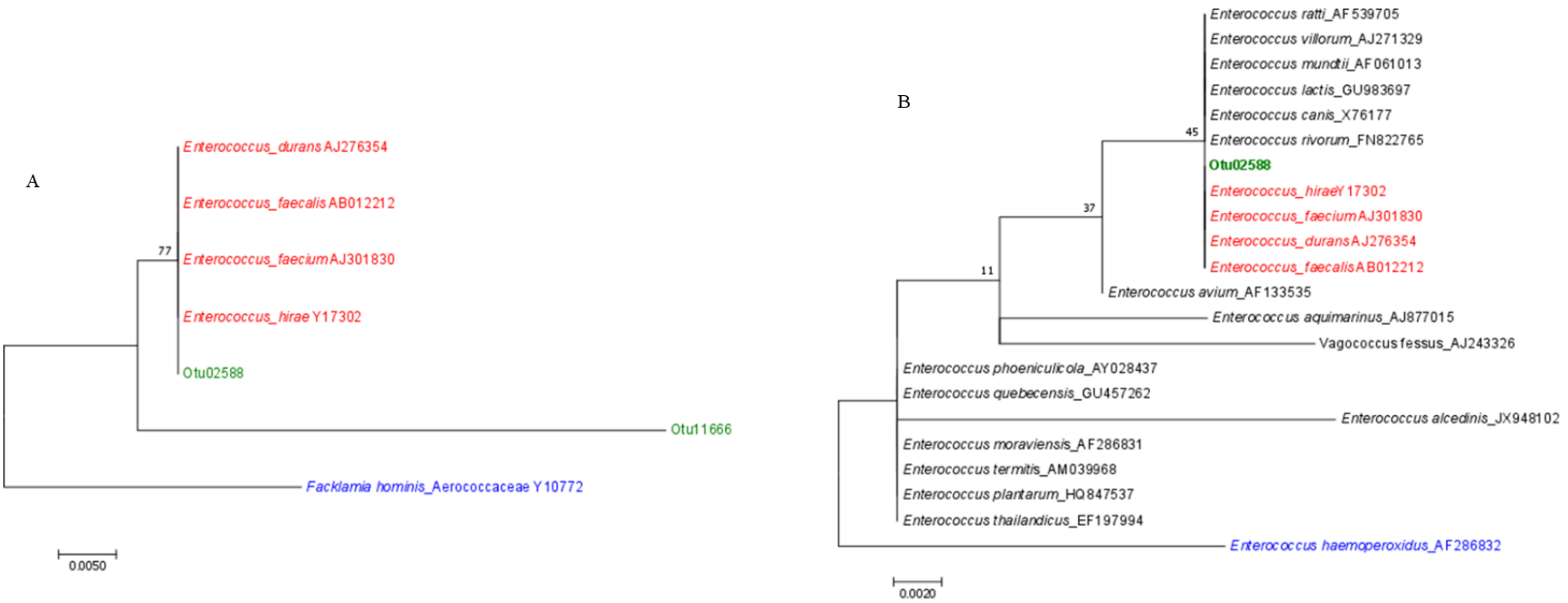


Figure 6.16 (A) Maximum likelihood tree representing potentially pathogenic OTUs within the family *Enterococcaceae*. (B) Maximum likelihood tree representing the potentially pathogenic OTU compared to both the potential pathogenic species as well as other non-pathogenic related species of the family *Enterococcaceae*. Values indicated are representative of 100 bootstrap replicates. The pathogenic bacterial strains are shown in red, OTUs classified up to the genus level (within the family) are shown in green and the outgroup is shown in blue.

6.3.3 Relative abundance

The relative abundance graphs for the 18 OTUs identified as potential pathogens by the phylogenetic analyses are represented in Figures 6.17-6.33. The relative abundances of the individual OTUs, as calculated across the 8 different sampling points across the respective months sampled, were plotted. No general trends, where any OTU dominated within the system during a specific month or at a specific sampling point, were observed. In most cases the OTUs were only present at relatively low abundances, but sporadic increases in abundance were noted at a specific sampling point. The relative abundance graph for OTU0754 is not included as the OTU was only observed twice in all the samples tested.

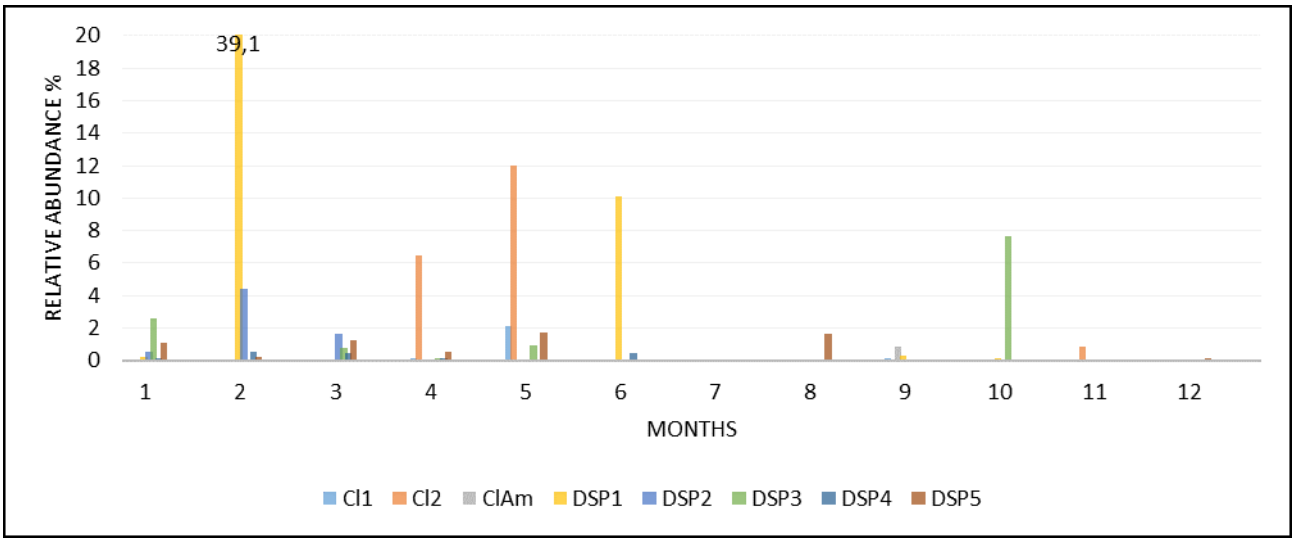


Figure 6.17 Relative abundance graph representing OTU00039 potentially *Novosphingobium aromaticivorans* across different months and sampling points.

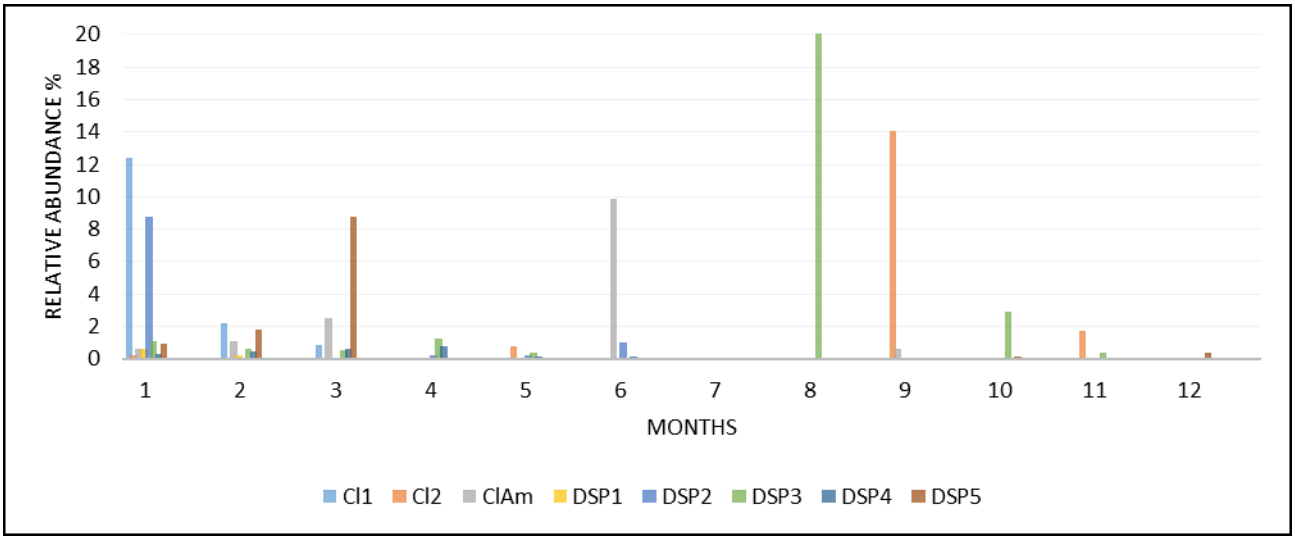


Figure 6.18 Relative abundance graph representing OTU00094 potentially *Aeromonas caviae* / *Aeromonas veronii* / *Aeromonas hydrophila subsp. hydrophila* / *Aeromonas jandaei* across different months and sampling points.

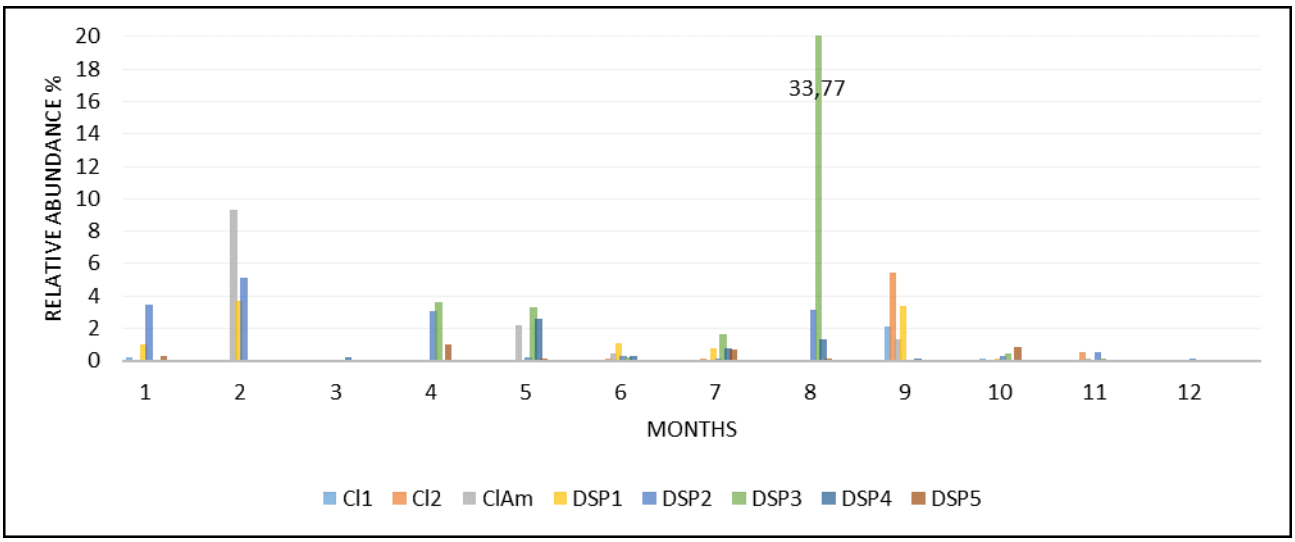


Figure 6.19 Relative abundance graph representing OTU00108 potentially *Staphylococcus warneri* / *Staphylococcus lugdunensis* / *Staphylococcus epidermidis* / *Staphylococcus capitis subsp. capitis* / *Staphylococcus aureus subsp. aureus* across different months and sampling points.

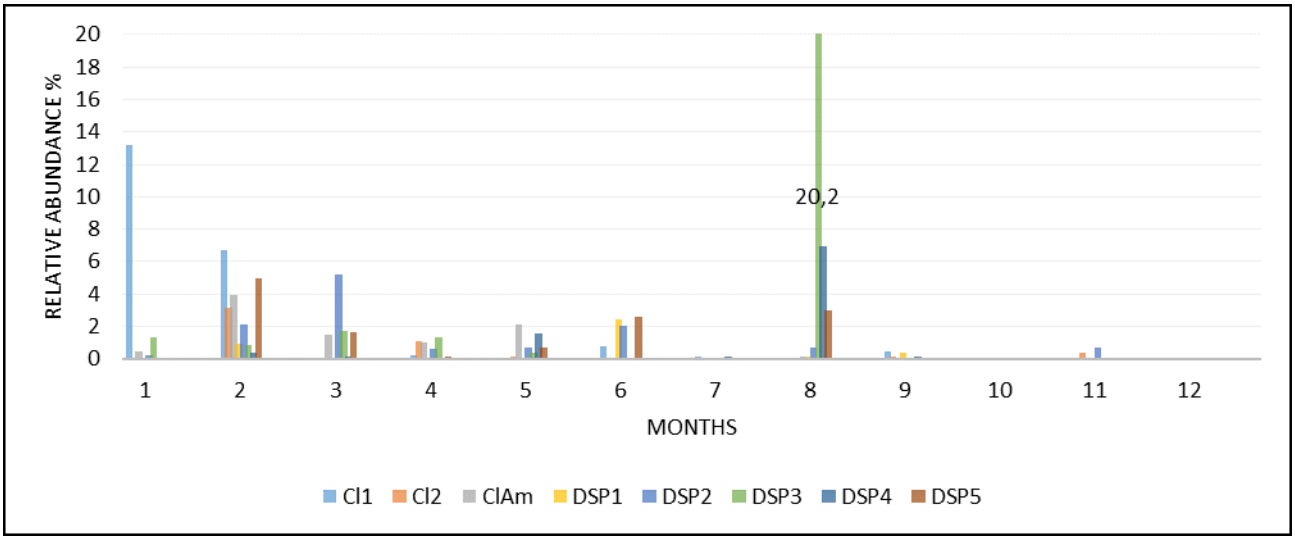


Figure 6.20 Relative abundance graph representing OTU00136 potentially *Escherichia coli* / *Escherichia fergusonii* / *Shigella sonnei* / *Shigella flexneri* across different months and sampling points.

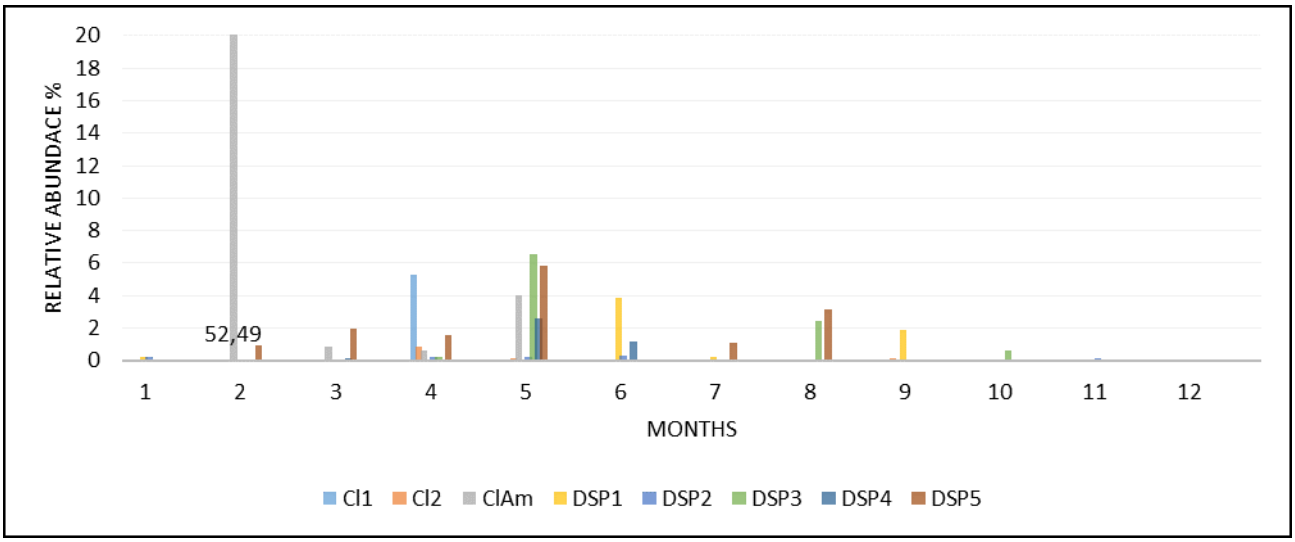


Figure 6.21 Relative abundance graph representing OTU00150 potentially *Methylobacterium radiotolerans* / *Methylobacterium mesophilicum* across different months and sampling points.

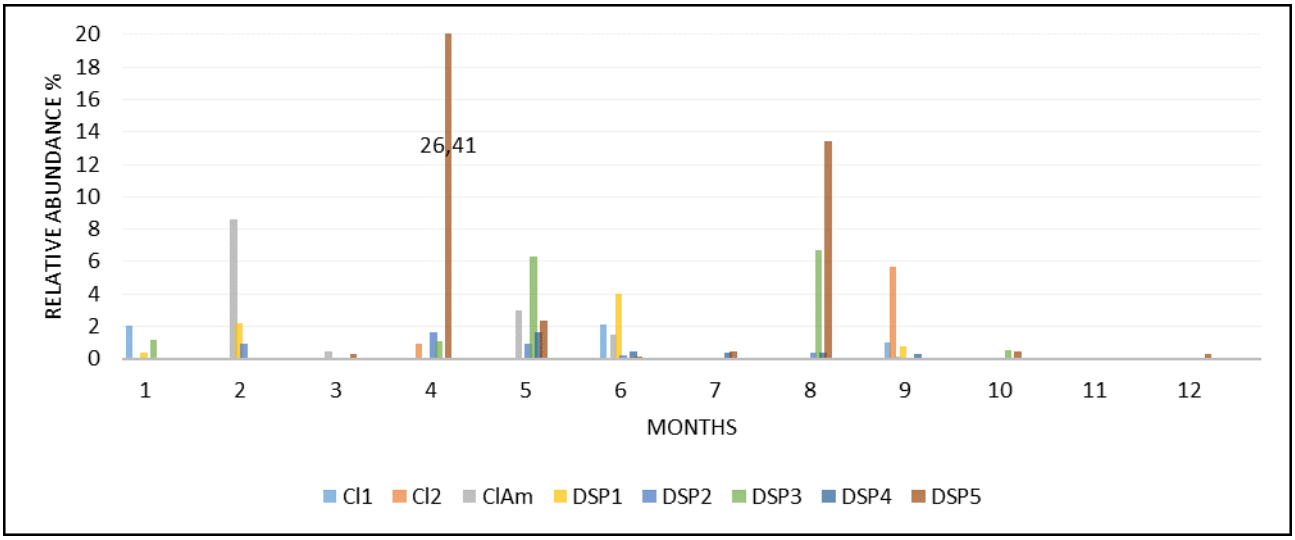


Figure 6.22 Relative abundance graph representing OTU00168 potentially *Streptococcus parasanguinis* across different months and sampling points.

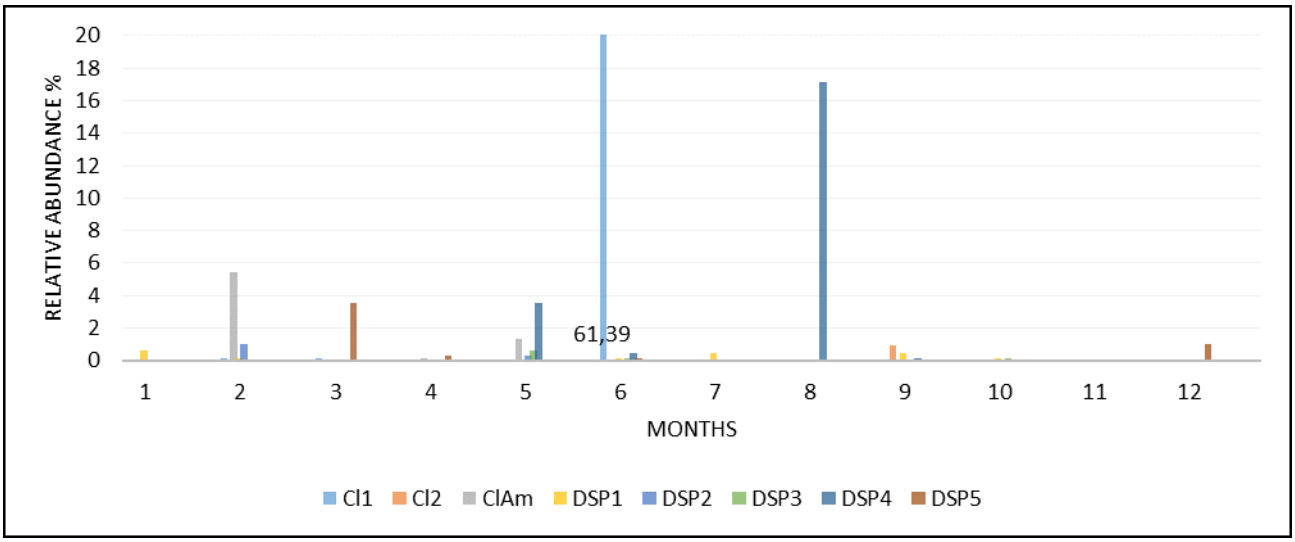


Figure 6.23 Relative abundance graph representing OTU00250 potentially *Klebsiella pneumoniae subsp. pneumoniae* / *Enterobacter aerogenes* / *Enterobacter asburiae* / *Enterobacter cancerogenus* / *Enterobacter hormaechei* / *Enterobacter intermedius* across different months and sampling points.

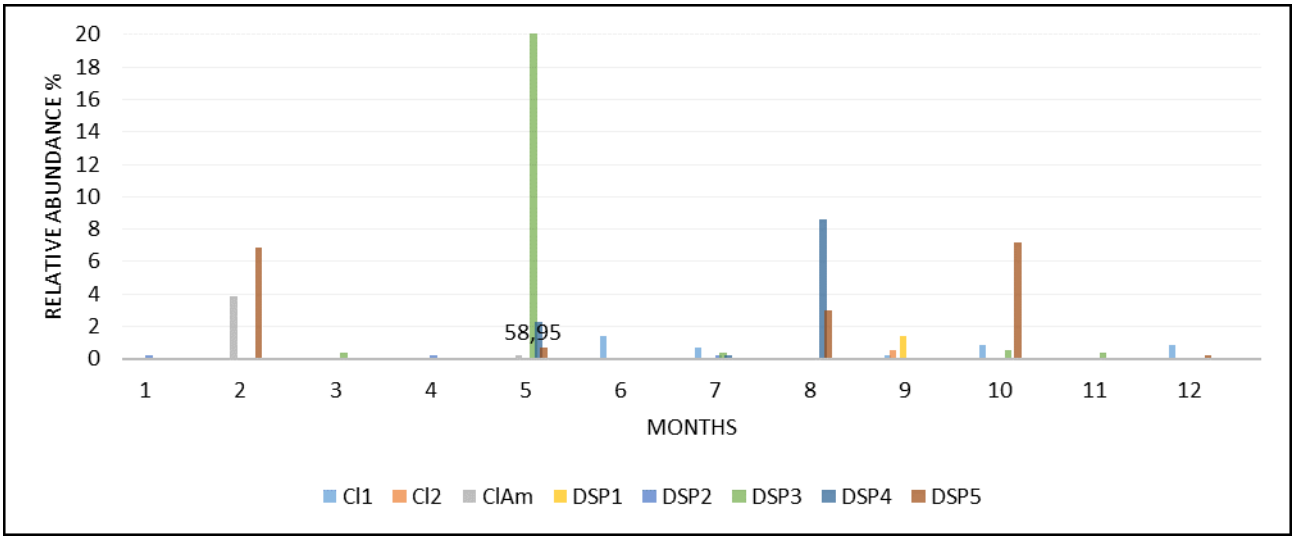


Figure 6.24 Relative abundance graph representing OTU00282 potentially *Moraxella osloensis* across different months and sampling points.

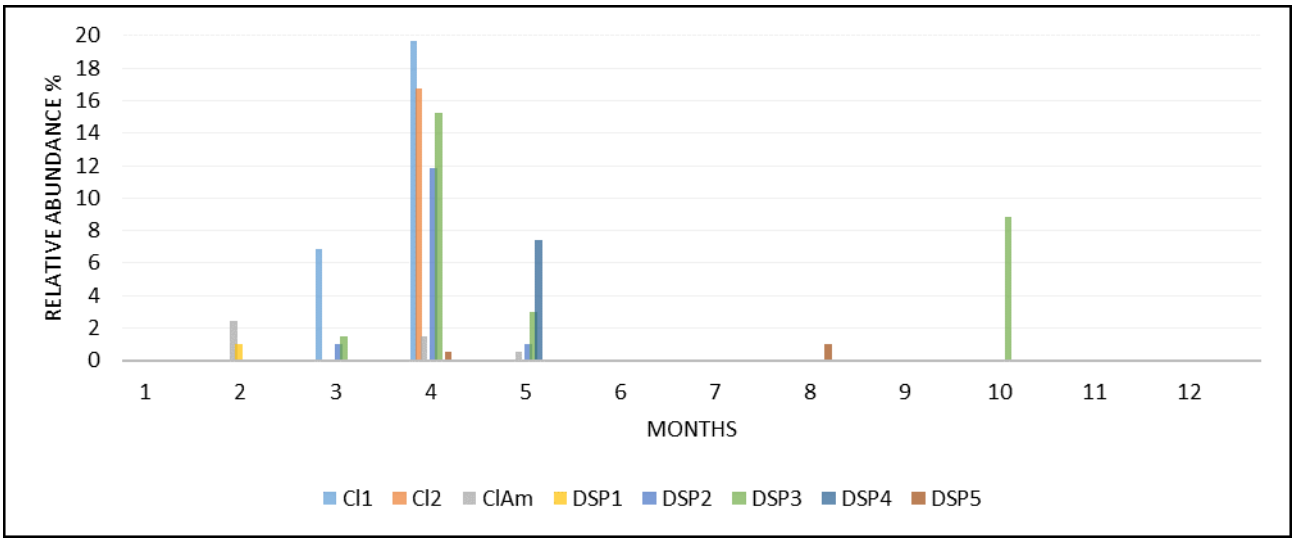


Figure 6.25 Relative abundance graph representing OTU00479 potentially *Methylobacterium aminovorans* across different months and sampling points.

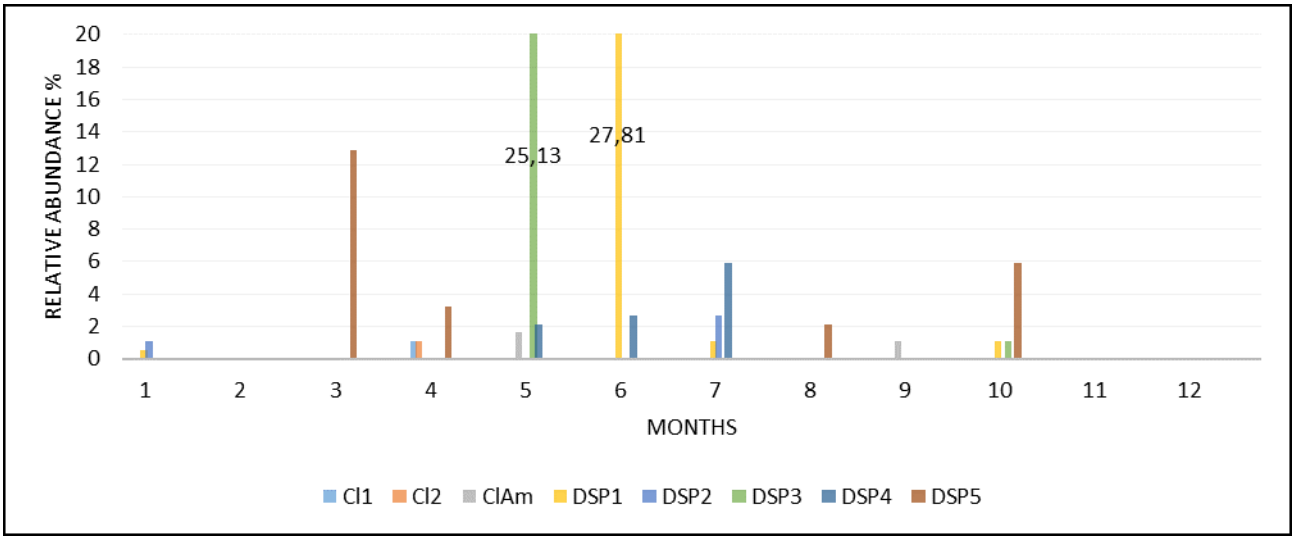


Figure 6.26 Relative abundance graph representing OTU00517 potentially *Corynebacterium tuberculostearicum* across different months and sampling points.

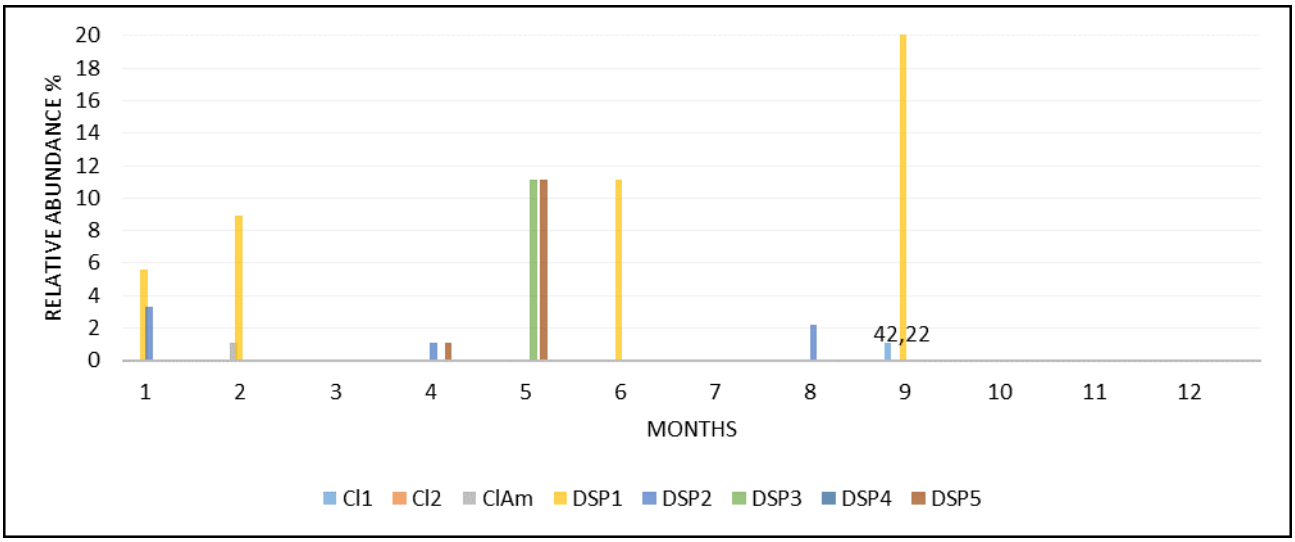


Figure 6.27 Relative abundance graph representing OTU00747 potentially *Micrococcus luteus* across different months and sampling points.

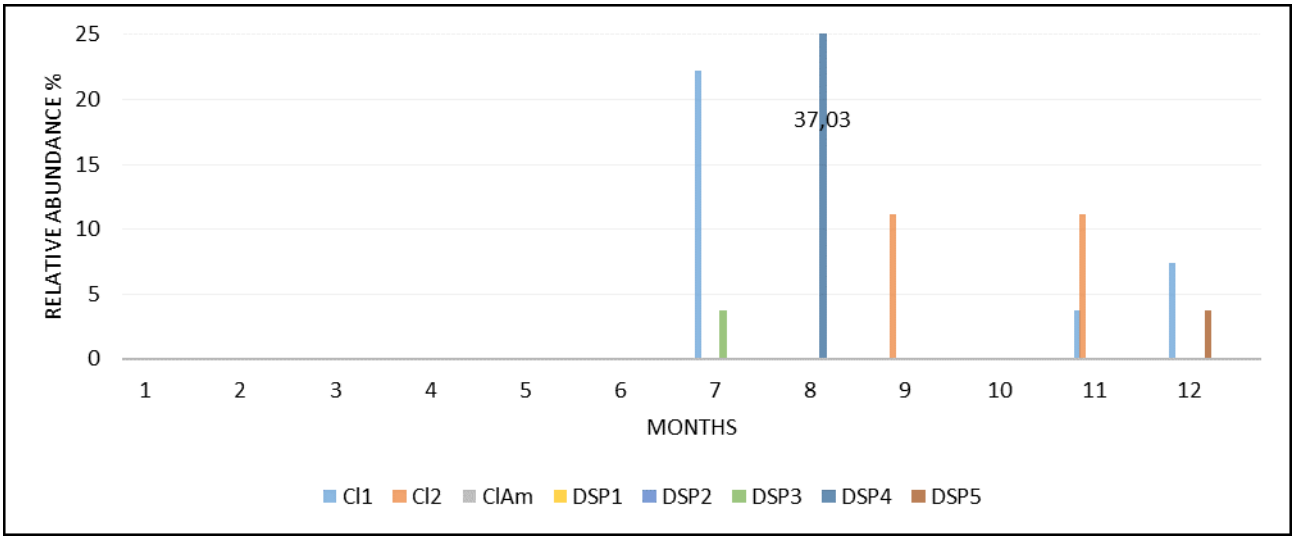


Figure 6.28 Relative abundance graph representing OTU01172 potentially *Burkholderia pseudomallei* across different months and sampling points.

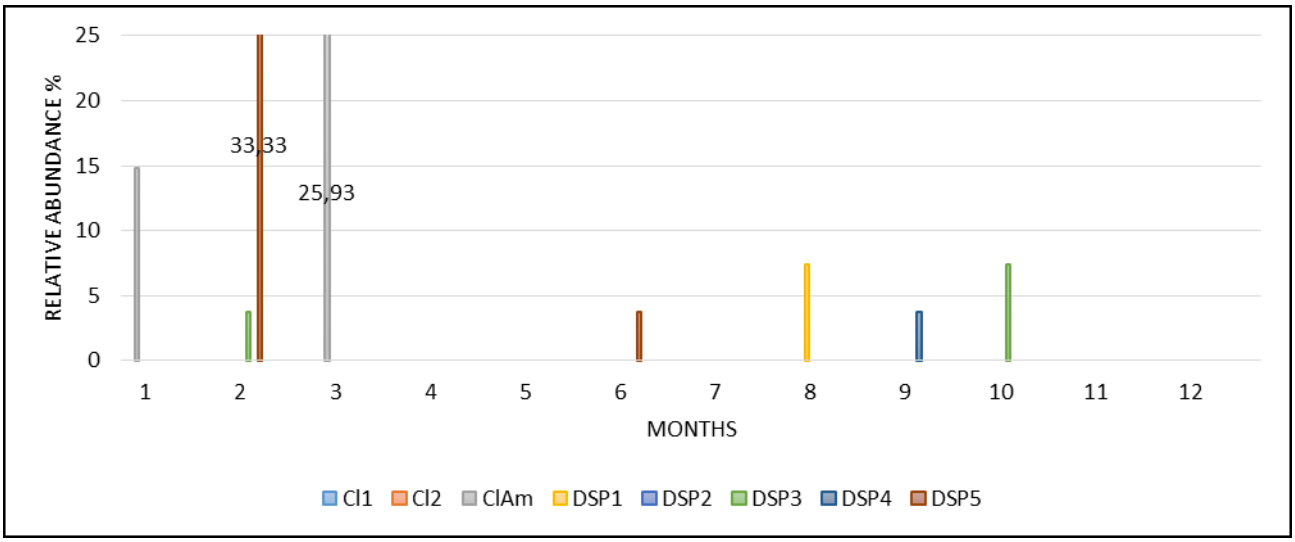


Figure 6.29 Relative abundance graph representing OTU01284 potentially *Yersinia enterocolitica subsp. enterocolitica* / *Serratia liquefaciens* / *Serratia plymuthica* across different months and sampling points.

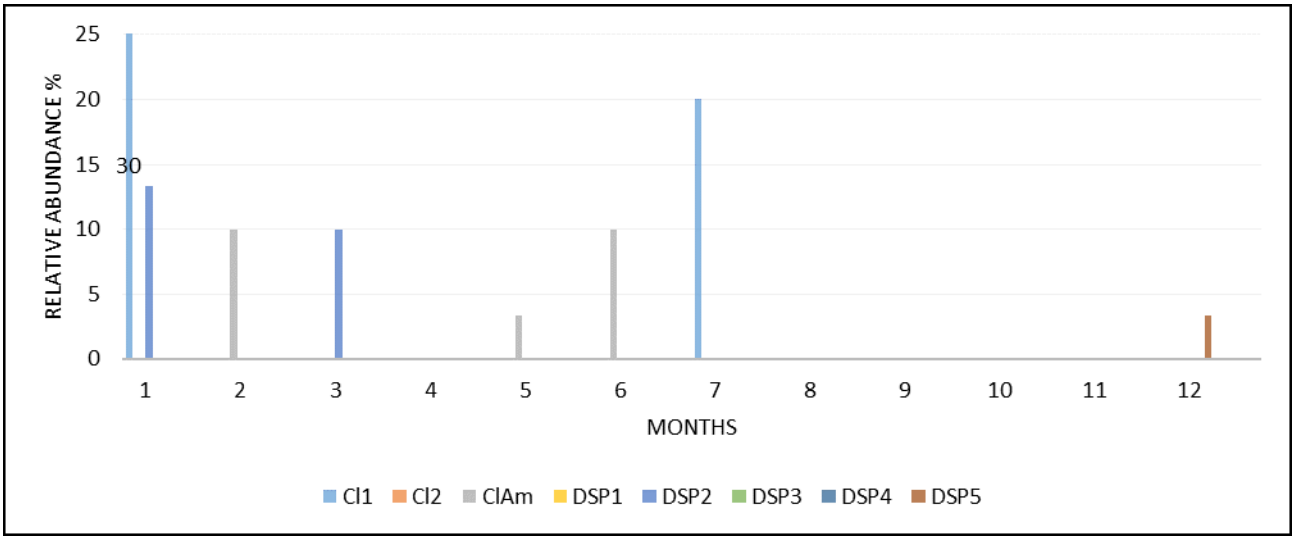


Figure 6.30 Relative abundance graph representing OTU01406 potentially *Mycobacterium lentiflavum* / *Mycobacterium genavense* / *Mycobacterium montefiorense* / *Mycobacterium simiae* across different months and points sampled.

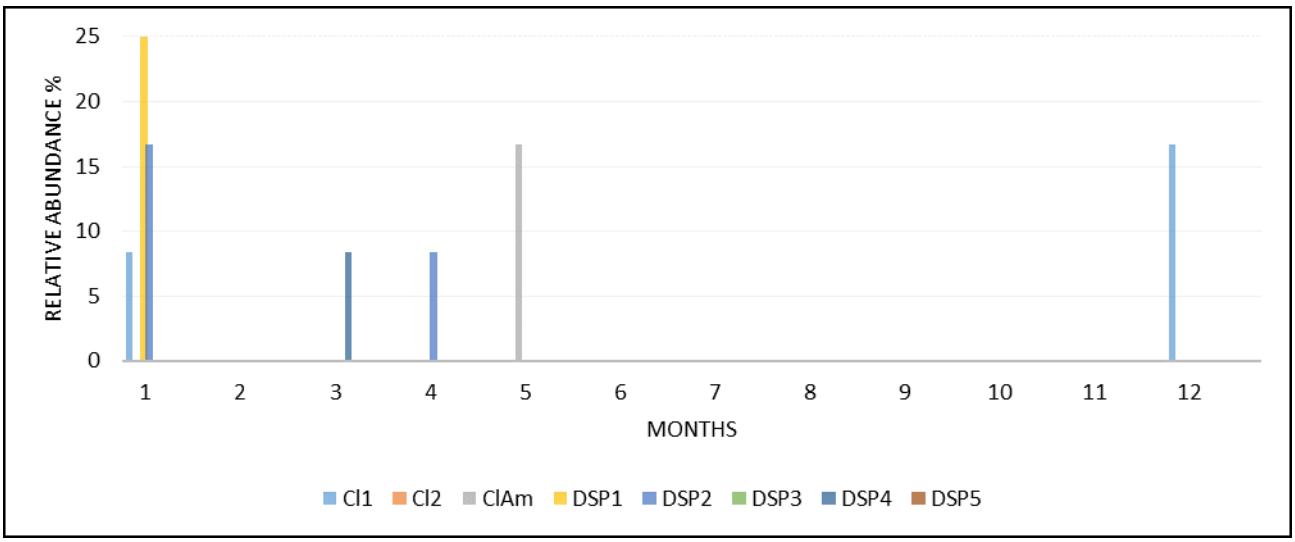


Figure 6.31 Relative abundance graph representing OTU01654 potentially *Bacillus cereus* across different months and sampling points.

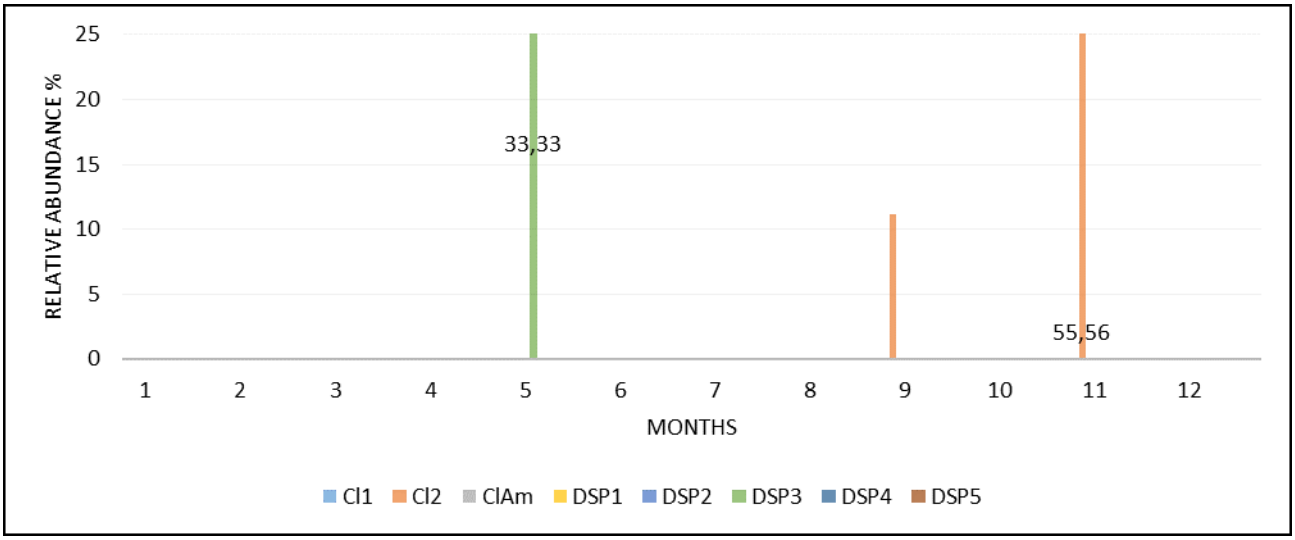


Figure 6.32 Relative abundance graph representing OTU01997 potentially *Alcaligenes faecalis subsp. faecalis* across different months and sampling points.

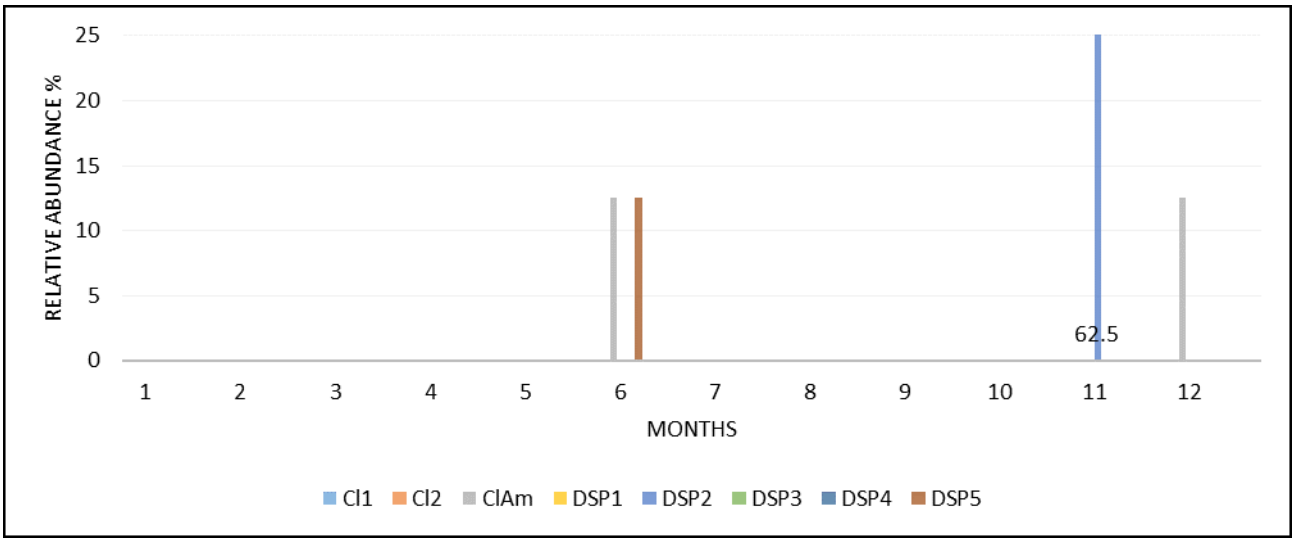


Figure 6.33 Relative abundance graph representing OTU02588 potentially *Enterococcus durans* / *Enterococcus faecalis* / *Enterococcus faecium* / *Enterococcus hirae* across different months and sampling points.

6.4 DISCUSSION

In this study the presence and abundance of potential bacterial pathogens, present in a large water distribution system in South Africa was determined. 16S profile sequence data representing the bacterial community present in the distribution system was used to identify OTUs that were closely related to the sequences of type strains representing potential waterborne and water-based opportunistic bacterial pathogens. This study provides an understanding of the potential bacterial pathogen community found in the distribution system across the sampling points studied over a period of one year.

Of the 18 OTUs which were closely associated with potential pathogens, eight were associated with more than one pathogenic strain due to the similarity (in some cases 100% e.g. Figure 6.3B and 6.4B) of the 16S rRNA V4 variable region sequence. The small sequence size (250 bp) used for the comparison is the main reason for the lack of resolution observed in the phylogenetic trees. The abundance of the majority of these OTUs were however low in relation to the original number of 8891 OTUs, which were observed across all the samples.

These 18 OTUs represented both waterborne and water-based environmental pathogens (opportunistic pathogens) commonly found in drinking water. Waterborne and water-based environmental pathogens are often confused as they often represent the same type of pathogen and are both transferred by water. They have the ability to cause diseases in humans, but based on unique characteristics related to their biology, these groups can broadly be distinguished from each other. The characteristics of each type of pathogen relate to the way in which they are introduced into the drinking water system, their ability to resist disinfection and treatment as well as the way in which they survive within the system. Enteric waterborne microorganisms typically come into contact with water sources primarily due to contamination from human or animal faecal matter (WHO, 2011); whereas the other group forms part of the natural community present in fresh water systems (Newton *et al*., 2011). These water-based environmental pathogens thrive in the water ecosystem as well as drinking water, which is suggested to be a vehicle for the spread of these pathogens (Van der Wielen and Van der Kooij, 2013). Several of the environmental pathogens are opportunistic pathogens which only affect susceptible individuals.

Of the OTUs classified as being pathogens, 14 OTUs were shown to be associated with water-based environmental pathogens (i.e. opportunistic) and 4 were associated with the group of waterborne pathogens. The presence of more potential water-based environmental pathogens is expected as these pathogens are naturally present in water ecosystems. The waterborne pathogens were probably introduced to the water system as part of the bacterial community present in the source water.

6.4.1 Waterborne pathogens

Two (OTU136 and OTU168) of the four waterborne pathogens were classified belonging to the family *Enterobacteriaceae*. OTU136 (potentially *Escherichia coli* / *Escherichia fergusonii* / *Shigella sonnei* / *Shigella flexneri* occurred frequently) was detected in the majority of the months sampled but mostly at very low abundance (Figure 6.20). The relative abundance was only high at two sampling points (Cl and DSP3) during different months i.e. month 1 and 8, respectively, in comparison to the relative abundance at other sampling points measured during these same months. The common occurrence of this OTU at a low relative abundance was expected as *Escherichia coli* is often associated with water. Due to its common occurrence in raw drinking water sources it is valued for its ability to be an indicator of treatment effectiveness (WHO, 2011). The pathogenic strains of *E. coli* are known to cause gastroenteritis and acute diarrhoea and pathogenic strains are additionally grouped based on clinical features etc. (Cabral, 2010). It is uncertain whether this OTU represented specific pathogenic strains.

OTU1284 (potentially *Yersinia enterocolitica subsp. Enterocolitica* / *Serratia liquefaciens* / *Serratia plymuthica*) was present across the majority of the months but not always present at every sampling point in that month (Figure 6.29). This bacteria is also sometimes used as an indicator of water quality. The occurrence of this potential pathogen at two different sampling points within the same month was shown to occur only once (month 2). Relatively higher abundances were observed on two instances at sampling points ClAm and DSP5. The presence of pathogenic strains of the bacteria in the environment are rare but can occur in untreated water (WHO, 2011).

OTU168 potentially identified as *Streptococcus parasanguinis* forming part of the family *Streptococcaceae*. It was detected across the majority of the sampling points but was present in relatively low abundances or in some cases not at all (Figure 6.22). The relative abundance was significantly high at sampling point DSP5 on two occasions (month 4 and 8) in relation to the abundance across all points and months sampled. The points at which the pathogen has been previously reported include drinking water taps. The pathogen is a normal inhabitant of the gastrointestinal tract and sometimes known to cause endocarditis in addition to bacteraemia and has also been shown to co-infect the urinary tract with *E*. *coli* (Felfoldi *et al*., 2009; Lee *et al*., 2014).

OTU2588 potentially *Enterococcus durans* / *Enterococcus faecalis* / *Enterococcus faecium* / *Enterococcus hirae* formed part of the family *Enterococcaceae.* This OTU was not highly abundant in the system and was shown to occur only four times across all the months sampled, occurring only twice during one month at different sampling points (Figure 6.33). There was a relatively high abundance during month 11 at DSP2 in comparison to the other three occurrences. The pathogenic abilities of this bacterium is not of great concern due to the sporadic relative abundance and low abundance in the system.

6.4.2 Water-based environmental pathogens (opportunistic pathogens)

Opportunistic pathogens are typically water-based microorganisms, which exist as part of the aquatic community and survive treatment (chlorination and chloramination) and as a result retain the ability to regrow

in the distribution system, posing a potential health risk to consumers. These bacteria are typically weakly virulent pathogens but have retained the ability of infecting immunocompromised humans to initiate disease (Van der Wielen and Van der Kooij, 2013). Based on previous studies, water-based environmental pathogens grow in the drinking water distribution systems as part of the community, either as part of biofilms attached to inner pipe walls or within amoeba hosts which live freely in the system (Ashbolt, 2015).

The potentially pathogenic OTU39 occurring in the family *Sphingomonadaceae* is *Novosphingobium aromaticivorans*. This OTU was generally found to occur sporadically at different sampling points in the months sampled but were not as frequently detected during in the warmer months towards the end of the sampling campaign. The OTU abundance at DSP1 in month 2 was however greater than the relative abundance at any of the other sampling points (Figure 6.17). This pathogen is a common inhabitant of water and was shown to trigger the onset of the autoimmune disease primary biliary cirrhosis (PBC), which infects the liver in addition to metabolizing xenobiotics and activating environmental estrogens (Padgett *et al*., 2005; Mohammed and Mattner, 2009; Mohammed *et al*., 2011; Rutebemberwa *et al*., 2014). The pathogen does however only infect individuals whom are genetically susceptible and enters the system through the digestive mucosa (Mohammed *et al*., 2011; Rutebemberwa *et al*., 2014). The infective dose for this pathogen to initiate infection is not known (Rutebemberwa *et al*., 2014).

The potentially pathogenic OTU94 was identified to be in the family *Aeromonadaceae* is *Aeromonas caviae*. The occurrence of this OTU was sporadic across all the sampling points and months. The abundance of the OTU was shown to be significantly different from the other months and was high during month 8 and 9 at two different sampling points (Figure 6.18). Based on other studies, this pathogen was shown to occur more commonly in the distribution system and originated from soil, sewage systems as well as in foods (Legnani *et al*., 1998; Lopes *et al*., 2015). The pathogen is known to commonly infect fish but may cause acute diarrhoea in humans when ingested due to the production of enterotoxins (Lopes *et al*., 2015).

The potentially pathogenic OTU108 was identified to be in the family *Staphylococcaceae* is *Staphylococcus warneri* / *Staphylococcus lugdunensis* / *Staphylococcus epidermidis* / *Staphylococcus capitis subsp. capitis* / *Staphylococcus aureus subsp. aureus*. The occurrence of the OTU is spread across all the months sampled (Figure 6.19) and a relatively high occurrence of the pathogen was observed at DSP3 during month 8. This was higher in comparison to the general occurrence across all the other sampling points and months. This might be due to increased levels of contamination in that month. Bacteria within this genus were shown to be ubiquitous in the environment and are frequent inhabitants of humans (WHO, 2011). Pathogenic species of this genus are often associated with infections such as bacteraemia and endocarditis as well as other soft-tissue infections (Bharadwaj and Sharma, 2016; WHO, 2011; Faria *et al*., 2009). Although this pathogen does inhabit drinking water (mostly occurring as part of biofilms), disease occurrence through consumption of water is not evident (WHO, 2011; Faria *et al*., 2009).

The potentially pathogenic OTU150 and OTU479 were identified to be in the family *Methylobacteriaceae* is *Methylobacterium radiotolerans* / *Methylobacterium mesophilicum* and *Methylobacterium aminovorans,*

respectively. The relative abundance of both OTUs across the respective months and sampling points was generally sporadic. OTU150 was present at ClAm during month 2, at significantly higher relative abundance than at the other points and was also shown to be rare in the later months (Figure 6.21). OTU479 was also seen to occur less frequently in the later months and occurred in considerably higher amounts during month 4 and 5 at respective sampling points than in other months (Figure 6.25). It could be that the growth of these OTUs was promoted by lower temperatures as they had a higher abundance during the winter months. *Methylobacterium mesophilicum* was reported to be the clinical strain most commonly found associated with nosocomial infection of immunocompromised patients (Rice *et al*., 2000). Pathogens from this genus were also found to be ubiquitous in drinking water (Vaz-Moreira *et al*., 2017). *Methylobacterium* also occurs as part of biofilms, grows slowly and resist disinfectants thereby retaining their ability to cause nosocomial infections. However, exposure to other sources such as ingestion of raw food, swimming in rivers and exposure to soil etc. can trigger infection in immunocompromised individuals (Kovaleva *et al*., 2014; Rice *et al*., 2000; Vaz-Moreira *et al*., 2017).

The potentially pathogenic OTU250 was identified to be in the family *Enterobacteriaceae* is *Klebsiella pneumoniae subsp. Pneumoniae* / *Enterobacter intermedius* / *Enterobacter hormaechei* / *Enterobacter cancerogenus* / *Enterobacter aerogenes* / *Enterobacter asburiae*. The OTU occurred rarely in the first and last few months and only showed a significant difference from the general relative abundance at two points (DSP2 and DSP4) during month 6 and 8, respectively (Figure 6.23). *Klebsiella* is different from the other pathogens in the *Enterobacteriaceae* family due to it being an opportunistic pathogen, which can cause pneumonia, nosocomial infection, septicaemia etc. but rarely causes infection in humans (Allen *et al*., 2004; Cabral, 2010; Geldreich, 2006; WHO, 2011). Pathogens in the *Enterobacter* genus generally cause urinary tract infections as well as bacteraemia (Cabral, 2010; WHO, 2011and 2014).

The potentially pathogenic OTU282 in the family *Moraxellaxeae* is *Moraxella osloensis*. The OTU's occurrence is sporadic and the relative abundance at point DSP3 during month 5 is significantly higher in comparison to the other months sampled (Figure 6.24). The OTU had a relatively low abundance in general across the months and sampling points. No gastrointestinal infections associated with this pathogen have been reported (WHO, 2011). The pathogen initiated infections such as septicaemia, meningitis, abscesses of various organs as well as infections and is needed in high amounts to cause infection which can also be transmitted via physical contact between people as well as from air and surfaces (in hospitals) (Stewart, 2006; WHO, 2011).

The potentially pathogenic OTU517 was identified to be in the family *Corynebacteriaceae* is *Corynebacterium tuberculostearicum*. The OTU occurred sporadically across the sampling points at the different months and showed increased relative abundance at points DSP3 and DSP1 during the colder months (month 5 and 6 respectively) in comparison to the other point's sampled (Figure 6.26). The pathogen is known to be associated with skin lesions and infections in mucosal surfaces (Felfoldi *et al*., 2009; Pindi *et al*., 2013).

The potentially pathogenic OTU747 was identified to be in the family *Micrococcaceae* is *Micrococcus luteus*. The OTU was shown to have a general high relative abundance at point DSP1 across several of the months sampled and occurred rarely or not at all during other months and at other sampling points (Figure 6.27). The pathogen is known to inhabit areas in the environment as well as the human body and is rarely known to cause infectious diseases (Jordaan, 2015).

The potentially pathogenic OTU754 was identified to be in the family *Bacteroidaceae* is *Bacteroides vulgatus*. A relative abundance graph for the OTU was constructed but not included because it showed the occurrence of the OTU at only two points during the entire sampling period. The pathogen does however, initiate inflammatory responses causing ulcerative colitis and inhabits the gut as part of the beneficial microflora. The pathogen may infect susceptible individuals opportunistically but evidence for infection through drinking water is inadequate (WHO, 2011; Falkinham *et al*., 2015; Jordaan, 2015).

The potentially pathogenic OTU1172 was identified to be in the family *Burkholderiaceae* is *Burkholderia pseudomallei*. The OTU was shown to occur only from month 7-12 (spring and summer) and was completely absent in the earlier months (Figure 6.28). The relative abundance differed significantly from one another at the various points at which it occurred, and was seen to occur mostly in the chlorine-treated water (Cl and BfAm) compared to its occurrence in the distribution system. The pathogen is known to cause melioidosis and is also the cause of septicemic pneumonia associated with drinking water (Ashbolt, 2004). The pathogen was also shown to inhabit the amoeba *Acanthamoeba astronyxis* increasing its survival through disinfection (Berry *et al*., 2006). Routes of infection other than ingestion of drinking water include inhalation. It causes infections in the brain, skin lesions as well as respiratory tract infections (WHO, 2011). The infective dose for this pathogen is not known (WHO, 2011).

The potentially pathogenic OTU1406 was identified to be in the family *Mycobacteriaceae* is *Mycobacterium simiae* / *Mycobacterium montefiorense* / *Mycobacterium lentiflavum* / *Mycobacterium genavense*. The OTU was generally seen to occur at mainly Cl, ClAm and DSP2 sampling points across the first seven months sampled and then was not detected during the remaining months (Figure 6.30). The reason for the OTU's occurrence at treatment points is most likely due to the pathogen being resistant to chlorination and chloramination. These pathogens are found in the aquatic environments and only cause disease in individuals whom are susceptible and have had prior infections. Majority of the mycobacterium pathogens are not as notorious as *Mycobacterium tuberculosis* is (Vaerewijck *et al*., 2005). The pathogens listed can however cause diseases of the skin, gastrointestinal tract and infections in HIV patients resulting in death in most cases (Vaerewijck *et al*., 2005; WHO, 2011; Van der Wielen *et al*., 2013). These pathogens can cause infection through inhalation, contact or ingestion of contaminated water sources (WHO, 2011).

The potentially pathogenic OTU1654 was identified to be in the family *Bacillaceae* is *Bacillus cereus*. The OTU was seen to occur across a few sampling points (Cl, ClAm, DSP2, DSP4) frequently and was detected only once during some months and was not present at all in the remaining months (Figure 6.31). The pathogen is known to cause food poisoning, which leads to diarrhoea and nausea and transmission through

drinking water has not been confirmed. Infections are commonly caused through the consumption of contaminated foods (Allen *et al*., 2004; WHO, 2011; Pindi *et al*., 2013).

The potentially pathogenic OTU1997 was identified to be in the family *Alcaligenaceae* is *Alcaligenes faecalis subsp. faecalis*. The OTU was seen to occur only three times across different months at either point BfAm or DSP3 and was not detected at all during the other months sampled (Figure 6.32). This pathogen is a common inhabitant of the natural environment and the pathogen is responsible for infections such as bacteraemia, septicaemia, skin and soft tissue infections etc. (Berry *et al*., 2006; Pindi *et al*., 2013). Evidence for gastrointestinal infection through drinking water was not found (Jordaan, 2015).

Flow cytometry data obtained from a previous study (Chapter 4) was also performed on samples collected from the mentioned sampling points. The data was collected for months 8-12 using approximately 50 ml of sampled water. The cell counts showed that very few of the cells remained intact from the total cell counts. The intact cells however are not an accurate representation of live cells with the capability to cause any sort of infection. The total number of intact cells in 1 ml of the sample varied between 4000 and zero at different sampling locations.

6.5 SUMMARY

In summary, the results showed that out of the 8891 OTUs, initially observed in the dataset, only eighteen OTUs were reported to be potential bacterial pathogens (waterborne and water-based- pportunistic) and eight of those OTUs were shown to be closely related to more than one pathogenic strain. The relative abundances of these respective OTUs together with the related pathogens were observed and shown to be of little concern due to their sporadic occurrences and low bacterial cell counts observed in the flow cytometry data. Most of the diseases associated with the different potential pathogens identified are also not significant due to the majority of the pathogens have other sources of being transmitted and causing infection other than drinking water. Cases reported due to drinking water consumption are not known. The presence of the waterborne pathogens were possibly due their survival of growth in the system and the presence of the water-based environmental OTUs were expected due to the ubiquitous nature of these pathogens in the environment. As mentioned previously, these pathogens have the ability to survive treatment processes and occur as part of the biofilm community within pipe walls (Van der Wielen and Van der Kooij, 2013). The pathogens detected however are not a major health risk to humans as the pathogens found to cause severe diseases in previous studies, such as *Legionella pneumophila* and *Mycobacterium avium complex,* were not found in this study. The infective dose for majority of these pathogens are not known and outbreaks associated with these pathogens have not been well documented. The health risk posed by the potential bacterial pathogens detected in the drinking water part of the distribution system in this study are of little concern as they occur sporadically and at generally low abundances in comparison to the total drinking water microbial community.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

7.1 CONCLUSIONS

It was previously reported that the composition of the bulk water community was consistently shaped by the dominant bacteria associated with the sand filter used during treatment (Pinto *et al.*, 2012; Pinto *et al.,* 2014). Pinto *et al*. (2014) therefore suggested that it may be possible that the bacterial community on the filter could be used to predict the bacterial communities downstream in the distribution system. One of the main objectives of the current study was therefore to determine the impact of the sand filter community on the bulk drinking water in the distribution system. A second objective, linked to the issue of developing a predictive system based on the sand filter community, was to determine whether the community within the sand filter is homogenous enough to use a single sample as representative of the whole community associated with the filter.

Contrary to what was initially expected, it was observed that in the large distribution system sampled during this project, sand filter community was not the main driver of the microbiome present the bulk water. This study showed that the microbial communities inhabiting the large drinking water distribution system were strongly shaped by the chlorination and subsequent chloramination that were administered to the water at different points within the system. This finding was not totally unexpected, as it is known that disinfection processes have a significant impact on the microbial community.

It is also known that multiple factors could affect the microbial growth in the bulk water especially in the large distribution systems found in South Africa. The interactions between communities associated with biofilms, loose deposits and the bulk water is also unclear. When considering the effects of hydraulic forces, the detachment of biofilms and re-suspension of sediments can undoubtedly contribute to bacterial community composition and cell concentrations in the bulk water (Prest *et al*., 2016).

Based on the sampling performed at two different treatment works, it was established that although some level of local and intra variation amongst the rapid sand filter microbial communities exist, no significant spatial differences were detected amongst the sand filter microbial communities when measured at different locations within the filter bed (i.e. along the surface and depth of the RS filter bed). The variability across parallel RS filters within the same filter house, sampled during the same sampling run, was also found to be insignificant. This data indicated that at any specific point in time, the sand filter community is homogenously distributed across the filter. This has definite implications for sampling efforts and is so far known to be the first study to address this issue in detail.

From the literature is was clear that information on the microbiome of large distribution systems would provide important information required for the management of the microbial water quality in these systems. The study also investigated whether South African communities supplied by smaller drinking water and reticulation systems could also benefit. This study demonstrated that a 16S community profiling approach could provide valuable information to better understand the impact of alternative treatment approaches or water sources on the microbial community present in the water supplied to consumers. For some of the of the smaller utilities and municipalities implementation of such a study would only be feasible when contracting experienced research organisation with the required skilled investigators and equipped facilities. The current average analysis cost of R500 or more per sample (depending on the number of samples analysed) is also high but it is predicted that the cost of these studies would become more affordable in the near future due to continuous technological advances in the field of DNA sequencing. In spite of these limitations, it is currently the only way in which a detailed understanding of the microbial community in drinking water systems can be obtained.

This study again confirmed that groundwater, from a well-protected aquifer, remains a valuable source of drinking water as microbes are present at very low levels. It was shown that for such systems only minimal disinfection and treatment are required to ensure safe drinking water. Attention should however, be given to the distribution of this water as no residual disinfectants are present to counter contamination events.

One of the main benefits of the 16S profiling approach is that it provides a detailed inventory of the bacterial species present in the drinking water distribution system. The sequence data does not only assist in determining the identity of the species present but could also be used to link them with known opportunistic pathogens. The data also provides an indication of their relative abundance at the time of sampling. Based on the data collected for the large distribution system it was felt that the health risk posed by the potential bacterial pathogens detected could be of limited concern as their distribution patterns observed were not indicative of either treatment failure or a specific contamination event.

This study has clearly shown that drinking water treatment and distribution systems could differ markedly from each other and that a universal model to predict the microbial community of the water supplied to the consumer would be difficult to achieve. The current study demonstrated that the necessary technologies and knowledge is available to study the microbial ecology and dynamics of individual treatment and distribution systems. Understanding the ecology and the factors that shape the drinking water microbiome is essential when appropriate measures to manage the microbial quality and associated health risks of drinking water in such a system, are to be developed and implemented.

7.2 RECOMMENDATIONS

This study clearly showed that each step within the distribution and reticulation system had an impact on the quality of the water that reaches the consumer. The main focus of the present study was on the distribution system and did not address the effects of the reticulation system on the microbial water quality in great detail. In reticulation systems, issues such as the impact of smaller diameter pipes well as retention times within the

smaller community reservoirs on the bacterial community are important. As these issues have not been addressed in detail during this study, it is recommended that future studies focusing specifically on the microbial ecology and dynamics of reticulation systems should be undertaken.

One of the shortcomings of the present study is that there is a strong focus on community membership but that information on the specific functions performed by the community is limited. This could be addressed by metagenome studies where the entire DNA in the sample is sequenced. The resulting sequence reads can then be used to identify protein coding sequences and predict their potential gene function and role in the metabolic processes performed by the community within the system. Furthermore, genomic binning of metagenomic data can be applied to investigate individual genomes of the dominant microorganisms in the system. This approach will clearly assist in understanding the biological interactions in the systems in greater detail and create valuable information, which could be used to manage the biostability of DWDSs.

LIST OF REFERENCES

Aizenberg-Gershtein, Y., Vaizel-Ohayon, D. and Halpern, M. (2012). Structure of bacterial communities in diverse freshwater habitats. *Canadian Journal of Microbiology*. 58: 326-335.

Alan, R., Liss, I., Shah, V.O. and Loken, M.R. (1988.) Discriminating between damaged and intact cells in fixed flow cytometric samples. *Applied Environmental.* 26: 484-490

Albers, C.N., Ellegaard-Jensen, L., Harder, C.B., Rosendahl, S., Knudsen, B.E., Ekelund, F. and Aamand, J. (2015). Groundwater chemistry determines the prokaryotic community structure of waterworks sand filters. *Environmental Science and Technology*. 49: 839-846.

Al-Bahrya, S.N., Al-Hinaib, J.A., Mahmoudc, I.Y. and Al-Musharafid, S.K. (2013). Opportunistic and Microbial Pathogens in Municipal Water Distribution Systems. *APCBEE Procedia.* 5: 5.

Allen, E.E. and Banfield, J.F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews*. 3: 489-498.

Allen, M.J., Edberg, S.C. and Reasoner, D.J. (2004). Heterotrophic plate count bacteria—what is their significance in drinking water? *International Journal of Food Microbiology.* 92: 10.

Ashbolt, N.J. (2003). WHO. Methods to identify and enumerate frank and opportunistic bacterial pathogens in water and biofilms. *Heterotrophic Plate Counts and Drinking-water Safety,* (J. Bartram, M. Exner, C. Fricker, A. Glasmacher, ed.) p. 31. IWA Publishing, London, UK.

Ashbolt, N.J. (2004). Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology.* 198: 10.

Ashbolt, N.J. (2015a). Environmental (Saprozoic) Pathogens of Engineered Water Systems: Understanding Their Ecology for Risk Assessment and Management. *OPEN ACCESS pathogens.* 4: 390-405.

Ashbolt, N.J. (2015b). Microbial Contamination of Drinking Water and Human Health from Community Water Systems. *Springer.* 2: 12.

Atlas, R.M. and Bartha, R. (1998). Microbial Ecology, 4th ed. Benjamin/Cummings, Menlo Park, CA, 694 pp.

Ayarza, J.M. and Erijman, L. (2011). Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. *Microbial ecology.* 61: 486-495.

Bai, Y., Liu, R., Liang, J. and Qu, J. (2013). Integrated metagenomic and physicochemical analyses to evaluate the potential role of microbes in the sand filter of a drinking water treatment system. *PLoS One*. 8: e61011.

Baron, J.L., Vikram, A., Duda, S., Stout, J.E. and Bibby, K. (2014). Shift in the microbial ecology of a hospital hot water system following the introduction of an on-site monochloramine disinfection system. *PLOS One.* 9(7): e102679.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O., Hunault, G., Simoneau, P. and Jacques, M.A. (2015). Emergence shapes the structure of the seed microbiota. *Applied and Environmental Microbiology*. 81: 1257-1266.

Bar-Zeev, E., Belkin, N., LibermaN, B., Berman, T. and Berman-Frank, I. (2012). Rapid sand filtration pretreatment for SWRO: Microbial maturation dynamics and filtration efficiency of organic matter. *Desalination.* 286: 120-130.

Batté, M., Appenzeller, B.M.R., Grandjean, D., Fass, S., Gauthier, V., Jorand, F., Mathieu, L., Boualam, M., Saby, S. and Block, J.C. (2003). Biofilms in drinking water distribution systems. *Reviews in Environmental Science and Biotechnology*. 2(2-4): 147-168.

Bautista-de los Santos, Q.M., Schroeder, M.C., Sevillano-Rivera, M.C., Sungthong, R., Ijaz, U.Z., Sloan, W.T. and Pinto, A.J. (2016). Emerging investigators series: microbial communities in full-scale drinking water distribution systems – a meta-analysis. *Environmental Science: Water Research and Technology.* doi: 10.1039/c6ew00030d.

Beech, I.B. and Sunner, J. (2004). Biocorrosion: towards understanding interactions between biofilms and metals. *Current Opinion in Biotechnology*. 15: 181-186.

Bell, T. (2010). Experimental tests of the bacterial distance-decay relationship. *The ISME Journal.* 4: 1357-1365.

Belser, L.W. (1976). Population Ecology of Nitrifying Bacteria. *Annual Review of Microbiology*. 33: 309-333.

Berger, P.S., Clark, R., Reasoner, D., Rice, E. and Santo Domingo, J. (2015). Water, Drinking. *Applied Microbiology: Agro/Food.* 17.

Berka, J., Chen, Y.-J., Leamon, J.H., Lefkowitz, S., Lohman, K.L., Makhijani, V.B., Rothberg, J.M., Sarkis, G.J., Srinivasan, M. and Weiner, M.P. (2010). Bead emulsion nucleic acid amplification. Google Patents.

Berney, M., Hammes, F., Bosshard, F., Weilenmann, H-U. and Egli, T. (2007). Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. *Applied and Environmental Microbiology* 73: 3283-3290.

Berney, M., Vital, M., Hulshoff, I., Weilenmann, H.U., Egli, T. and Hammes, F. (2008). Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water Research.* 42(14): 4010-4018.

Berry, D., Xi, C. and Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Current Opinion Biotechnology.* 17: 297-302.

Biers, E.J., Sun, S. and Howard, E.C. (2009). Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Applied and environmental microbiology.* 75: 2221-2229.

Bharadwaj, N.D.A. and Sharma, A.K. (2016). Detection of Escherichia coli, Staphylococcus Aureus and Salmonella Typhi in Drinking Water of Government Institutions and Organizations of Gwalior City. *International Journal of Engineering Sciences and Research Technology.* 5: 6.

Bitton, G. (2014). Microbiology of drinking water production and distribution, John Wiley and Sons. Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F. and Kjelleberg, S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology*. 73: 278-288.

Boe-Hansen, R., Albrechtsen, H.J., Arvin, E. and JØrgensen, C. (2002). Bulk water phase and biofilm growth in drinking water at low nutrient concentrations. *Water Research.* 36: 4477-4486.

Bougeard, C.M.M., Goslan, E.H., Jefferson, B. and Parson, S.A. (2010). Comparison of the disinfection by-product formation potential of treated waters exposed to chlorine and monochloramine. *Water Research.* 44: 729-740.

Brettar, I. and Hofle, G.M. (2008). Molecular assessment of bacterial pathogens —a contribution to drinking water safety. *ELSEVIER.* 19: 7.

Cabral, J.P.S. (2010). Water Microbiology. Bacterial Pathogens and Water. *International Journal of Environmental Research and Public Health.* 7: 47.

Camper, A.K. (2014). Organic matter, pipe material, disinfectants and biofilms in distribution systems. In *Microbial Growth in Drinking Water Supplies: Problems, Causes, Control and Research needs.* Van der Kooij, D and Van der Wielen, P.W.J.J. (eds). London, UK: IWA Publishing. pp. 77-94.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., *et al*. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*. 6: 1621-1624.

Cerrato, J.M., Falkinham, J.O., Dietrich, A.M., Knocke, W.R., Mckinney, C.W. and Pruden, A. (2010). Manganese-oxidizing and-reducing microorganisms isolated from biofilms in chlorinated drinking water systems. *Water Research.* 44: 3935-3945.

Chandran, A., Vargheseb, S., Kandelerc, E., Thomas, A., Hathad, M. and Mazumdera, A. (2011). An assessment of potential public health risk associated with the extended survival of indicator of pathogenic bacteria in freshwater lake sediments. *ELSEVIER.* 214: 7.

Chandy, J.P. and Angles, M.L. (2001). Determination of nutrients limiting biofilm formation and the subsequent impact on disinfectant decay. *Water Research.* 35(11): 2677-2682.

Chao, Y., Mao, Y., Wang, Z. and Zhang, T. (2015). Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing. *Scientific Reports.* 5: 10044. DOI: 10.1038/srep10044.

Chiao, T., Clancy, T.M., Pinto, A., Xi, C. and Raskin, L. (2014). Differential resistance of drinking water bacterial populations to monochloramine disinfection. *Environmental Science and Technology.* 48: 4038-3-37.

Chowdhury, S. (2012). Heterotrophic bacteria in drinking water distribution system: a review. *Environmental Monitoring and Assessment.* 184: 6087-6137.

Chu. C., Lu. C. and Lee. C. (2005). Effects of inorganic nutrients on the regrowth of heterotrophic bacteria in drinking water distribution systems. *Journal of Environmental Management.* 74: 255-263.

Cunliffe, D.A. (1991). Bacterial nitrification in chloraminated water supplies. *Applied and Environmental Microbiology*. 57(11): 3399-3402.

Davies, P.D. and Wheatley, A.D. (2012). Pilot plant study of alternative filter media for rapid gravity filtration. *Water Science and Technology.* 66: 2779-2784.

De Mandal, S., Kumari Panda, A., Singh Bisht, S. and Nachimuthu Senthil Kumar, N. (2015). Microbial Ecology in the Era of Next Generation Sequencing. *Next Generation: Sequencing and Applications.* S1: 6.

De Vet, W., Dinkla, I., Muyzer, G., Rietveld, L. and Van Loosdrecht, M. (2009). Molecular characterization of microbial populations in groundwater sources and sand filters for drinking water production. *Water Research*. 43: 182-194.

Donlan, R.M. and Pipes, W.O. (1988). Selected Drinking Water Characteristics and Attached Microbial Population Density. *Journal AWWA.* 80(11): 70-76.

Douterelo, I., Husband, S. and Boxall, J.B. (2014). The bacterial composition of biomass recovered by flushing and operational drinking water distribution system. *Water Research.* 54: 100-114.

Douterelo, I., Boxall, J.B, Deines, P., Sekar, R., Fish, K.E. and Biggs, C.A. (2014). Methodological approaches for studying the microbial ecology of drinking water distribution systems. *ELSEVIER.* 23.

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: 2194-2200.

Edwards, U., Rogall, T., Blöcker, H., Emde, M. and Böttger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic acids research.* 17: 7843-7853.

Emtiazi, F., Schwartz, T., Marten, S.M., Krolla-Sidenstein, P. and Obst, U. (2004). Investigation of natural biofilms formed during the production of drinking water from surface water embankment filtration. *Water Research.* 38: 1197-1206.

Edzwald, J.K. (1993). Coagulation in drinking water treatment: particles, organics and coagulants. *Water Science and Technology.* 27(11): 21-35.

Edzwald, J.K. (2010). Water quality and treatment a handbook on drinking water, McGraw Hill. Eichler, S., Christen, R., Höltje, C., Westphal, P., Bötel, J., Brettar, I., Mehling, A and Höfle, M. G. (2006). Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA-and DNA-based 16S rRNA gene fingerprinting. *Applied and Environmental Microbiology*. 72: 1858-1872.

Eichler, S., Christen, R., Holtje, C., Westphal, P., Botel, J., Brettar, I., Mehling, A. and Höfle, M.G. (2006). Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Applied and Environmental Microbiology.* 72(3): 1858-1872.

Eiler, A. and Bertilsson, S. (2007). *Flavobacteria* blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. *Applied and environmental microbiology.* 73: 3511-3518.

El-Chakhtoura, J., Prest, E., Saikaly, P., Van Loosdrecht., H.F. and Vrouwenvedler, H. (2015). Dynamics of bacterial communities before and after distribution in a full-scale drinking water network. *Water Research.* 74: 180-190.

Emde, K.M. E., Smith, D.W. and Facey, R. (1992). Initial investigation of microbially influenced corrosion (MIC) in a low temperature water distribution system. *Water Research.* 26(2): 169-175.

Ewerts, H. (2010). *Effectiveness of purification processes in removing algae from Vaal Dam water at the Rand Water Zuikerbosch treatment plant in Vereeniging.* North-West University.

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 131: 479-491.

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M. and Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. 2(6): 1-7.

Falkinham, J.O. (2016). Current Epidemiologic Trends of the Nontuberculous Mycobacteria (NTM). *Topical Collection on Water and Health.* 7.

Falkinham, J.O., Pruden, A. and Edwards, M. (2015). Opportunistic Premise Plumbing Pathogens: Increasingly Important Pathogens in Drinking Water. *OPEN ACCESS pathogens.* 4: 14.

Faria, C., Vaz-Moreira, I., Serapicos, E., Nunes, O.C. and Manaia, C.M. (2009). Antibiotic resistance in coagulase negative staphylococci isolated from wastewater and drinking water. *Science of the Total Environment*. 407: 7.

Farkas, A., Dragan-Bularda, M., Muntean, V., Ciataras, D. and Tigan, S. (2013). Microbial activity in drinking water-associated biofilms. *Central European Journal of Biology.* 8(2): 201-214.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. and Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*. 34: e22-e22.

Felfoldi, T., Heeger, Z., Vargh, M.A. and Marialigeti, K. (2009). Detection of potentially pathogenic bacteria in the drinking water distribution system of a hospital in Hungary. *European Society of Clinical Microbiology and Infectious Diseases.* 16: 4.

Feng, S., Chen, C., Wang, Q., Zhang, X., Yang, Z. and Xie, S. (2013). Characterization of microbial communities in a granular activated carbon-sand dual media filter for drinking water treatment. *International Journal of Environmental Science and Technology*. 10: 917-922.

Feng, S., Xie, S., Zhang, X., Yang, Z., Ding, W., Liao, X., Liu, Y. and Chen, C. (2012). Ammonium removal pathways and microbial community in GAC-sand dual media filter in drinking water treatment. *Journal of Environmental Sciences.* 24: 1587-1593.

Fernandez, G., Richter, M., Schuler, M., Pinhassi, J., Acinas, S.G., Gonzalez, J.M. and Pedros, A.C. (2013). Ecology of marine Bacteroidetes: a comparative genomics approach. *Publications in* Medicine. 5: 1026-1037

Ferrenberg, S., O'neill, S.P., Knelman, J.E., Todd, B., Duggan, S., Bradley, D., Robinson, T., Schmidt, S.K., Townsend, A.R. and Williams, M.W. (2013). Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *The ISME Journal*. 7: 1102-1111.

Flemming, H-C., Percival, S.I. and Walker, J.T. (2002). Contamination potential of biofilms in water distribution systems. *Water Supply.* 2: 271-280.

Flemming, H-C., Bendinger, B., Exner, M., Gebel, J., Kistemann, T., Schaule, G., Szewzyk, U. and Wingender, J. (2014). The Last Meters before the Tap: Where Drinking. In: *Microbial Growth in Drinking-Water Supplies – Problems, Causes, Control and Research Needs*. Van der Kooij, D and Van der Wielen, P. W. J. J. (eds). London, UK: IWA Publishing. pp: 207-238.

Fonseca, A.C., Summers, R.S. and Hernandez, M.T. (2001). Comparative measurements of microbial activity in drinking water biofilters. *Water Research.* 35: 3817-3824.

Foster, J.A., Bunge, J., Gilbert, J.A. and Moore, J.H. (2012). Measuring the microbiome: perspectives on advances in DNA-based techniques for exploring microbial life. *Briefings in Bioinformatics*. bbr080.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E. and Oakley, B.B. (2005). Ubiquity and diversity of ammonia-oxidising archaea in water columns and sediments of the ocean. *PNAS*. 102(41): 14683-14688.

Francois, T., Hendrik, H., Etienne, R., Mirjam, C. and Gurvan, M. (2011). Environmental and gut Bacteroidetes: the food connection. *Frontiers in Microbiology.* 2: 93-94

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. and Naeem, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proceedings of the National Academy of Sciences.* 103: 13104-13109.

Fuhrman, J.A., Cram, J.A. and Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews. Microbiology.* 13: 133.

Gall, A.M., Mariñas, B.J., Lu, Y. and Shisler, J.L. (2015). Waterborne Viruses: A Barrier to Safe Drinking Water. *PLOS Pathogens.* 11(6): e1004867. doi: 10.1371/journal.ppat.1004867.

Garcia, S.L., Mcmahon, K.D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., Woyke, T. and Warnecke, F. (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal.* 7: 137-147.

Garrity, G.M.J., Bell, J.A. and Liburn, T.G. (2003). Taxonomic outline of the prokaryotes. *Systematic* Bacteriology. 4: 1-397.

Geldreich, E.E. (1996). Microbial quality of water supply in distribution systems. CRC Press, Boca Raton, Fla.

Geldreich, E.E. (2006) Klebsiella. p. 4. American Water Works Association.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A. and Rodriguez-Valera, F. (2014). Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Molecular ecology.* 23: 6073-6090.

Gibert, O., Lefèvre, B., Fernández, M., Bernat, X., Paraira, M., Calderer, M. and Martínez-Lladó, X. (2013). Characterising biofilm development on granular activated carbon used for drinking water production. *Water Research*. 47: 1101-1110.

Gillespie, S., Lipphaus, P., Green, J., Parsons, S., Weir, P., Juskowiak, K., Jefferson, B., Jarvis, P. and Nocker, A. (2014). Assessing microbiological water quality in drinking water distribution systems with disinfectant residual using flow cytometry. *Water Research.* 65: 224-234.

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*. 11: 759-769.

Gobet, A., Quince, C. and Ramette, A. (2010). Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. *Nucleic Acids Research.* 38: e155-e155.

Gomez-Alvarez, V., Revetta, R.P. and Santo Domingo, J.W. (2012). Metagenomic analysis of drinking water receiving different disinfection treatments. *Applied and Environmental Microbiology*. 78(17): 6095-6102.

Gomez-Smith, C.K., LaPara, T.M. and Hozalski, R.M. (2015). Sulfate reducing bacteria and mycobacteria dominate the biofilm communities in a chloraminated drinking water distribution system. *Environmental Science and Technology.* 49: 8432-8440.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., Del Campo, J., Escudero, L., Rodríguez-Martínez, R. and Alonso-Sáez, L. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (*Flavobacteria*). *Proceedings of the National Academy of Sciences.* 105: 8724-8729.

Goslan, E.H., Krasner, S.W., Bower, M., Rocks, S.A., Holmes, P., Levy, L.S. and Parsons, S.A. (2009). A comparison of disinfection by-products found in chlorinated and chloraminated drinking waters in Scotland. *Water Research.* 43: 4698-4706.

Gülay, A., Musovic, S., Albrechtsen, H.J., Al-Soud, W.A., Sorensen, S.J. and Smets, B.F. (2016). Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. *The ISME Journal.* 16. doi:10.1038/ismej.2016.

Gulay, A. and Smets, B.F. 2015. An improved method to set significance thresholds for β diversity testing in microbial community comparisons. *Environmental microbiology,* 17, 3154-3167.

Hallam, N.B., West, J.R., Forester, C.F. and Simms, J. (2001). The potential for biofilm growth in water distribution systems. *Water Research.* 35(17): 4063-4071.

Hammes, F., Berney, M., Wang, Y., Vital, M., Koster, O. and Egli, T. (2008). Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Research.* 44(17): 4868-4877.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. and Martiny, J.B. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology.* 10: 497-506.

Henne, K., Kahisch, L., Brettar, I. and Holfe, M.G. (2012). Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany. *Applied and Environmental Microbiology.* 78(10): 3530-3538.

Hodgson, K. and Manus, L. (2006). A drinking water quality framework for South Africa. 6.

Hoefel, D., Monis, P.T., Grooby, W.L., Andrews, S. and Saint, C.P. (2005). Culture-independent techniques for rapid detection of bacteria associated with the loss of chloramine residual in a drinking water system. Applied and Environmental Microbiology. 71(11): 6479-6488.

Hoefel, D., Monis, P., Grooby, W., Andrews, S. and Saint, C. (2005). Profiling bacterial survival through a water treatment process and subsequent distribution system. Journal of applied microbiology. 99: 175-186.

Holinger, E.P., Ross, K.A., Robertson, C.E., Stevens, M.J., Harris, J.K. and Pace, N.R. (2014). Molecular analysis of point-of-use municipal drinking water microbiology. Water research 49: 225-235.

Hong, P., Hwang, C., Ling, F., Anderson, G.L., LeChevallier and Lui, W. (2010). Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. Applied and Environmental Microbiology. 76(16): 5631-5635.

Hong, C., Si, Y., Xing, Y. and Li, Y. (2015). Illumina MiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas. Environmental Science and Pollution Research. 22: 10788-10799.

Hubert, C., Loy, A., Nickel, M., Arnosti, C., Baranyi, C., Brüchert, V., Ferdelman, T., Finster, K., Christensen, F.M. and De Rezende, J.R. (2009). A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. Science. 325: 1541-1544.

Hunter, P.R., MacDonald, A.M. and Carter, R.C. (2010). Water supply and health. PLOS. Medicine. 7(11): e1000461.

Huisman, L., Wood, W. and WHO. (1974). Slow sand filtration. World Health Organization Geneva.

Hull, N.M., Holinger, E.P., Ross, K.A., Robertson, C.E., Harris, J.K., Stevens, M.J. and Pace, N.R. (2017). Longitudinal and Source-to-Tap New Orleans, LA, USA Drinking Water Microbiology. Environmental Science and Technology. 51: 4220-4229.

Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W. and Liu, W. (2012a). Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. Microbes and Environments. 27(1): 9-18.

Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W. and Liu, W. (2012b). Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. Applied and Environmental Microbiology. 78(22): 7856-7865.

Inkinen, J., Kaunisto, T., Pursiainen, A., Miettinen, I.T., Kusnetsoc, J., Riihinen, K. and Keinänen-Toivola, M.M. (2014). Drinking water quality and formation of biofilms in an office building during its first year of operation, a full scale study. Water research. 49: 83-91.

Ives, K.J. (1970). Rapid filtration. Water Research. 4: 201-223.

Ivone, V.M., Conceição, E., Olga, C.N. and Célia, M.M. (2013). Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. FEMS Microbiology Ecology. 83: 361-374.

Jordaan, K. (2015). Diversity and prevalence of potential opportunistic pathogens associated with one of the Rand Water distribution networks. p. 133. University of Pretoria, Faculty of Natural and Agricultural sciences.

Junli, H., Li, W., Nanqi, R, Fang, M. and Juli. (1997). Disinfection effect of chlorine dioxide on bacteria in water. Water Research. 31(3): 607-613.

Kahlisch, L., Henne, K., Gröbe, L., Brettar, I. and Höfle, M.G. (2012). Assessing the viability of bacterial species in drinking water by combined cellular and molecular analyses. Microbial Ecology. 63: 383-397.

Kalmbach, S., Manz, W. and Szewzyk, U. (2006). Dynamics of biofilm formation in drinking water: phylogenetic affiliation and metabolic potential of single cells assessed by formazan reduction and in situ hybridization. FEMS Microbiology Ecology. 22: 15.

Kanagal-Shamanna, R. (2016). Emulsion PCR: Techniques and Applications. Clinical Applications of PCR. 33-42.

Kasuga, I., Shimazaki, D. and Kunikane, S. (2007). Influence of backwashing on the microbial community in a biofilm developed on biological activated carbon used in a drinking water treatment plant. Water Science and Technology. 55: 173-180.

Keinänen-Toivola, M.M., Revetta, R.P. and Santo Domingo, J.W. (2006). Identification of active bacterial communities in a model drinking water biofilm system using 16S rRNA-based clone libraries. FEMS microbiology letters. 257: 182-188.

Kim, T.G., Yun, J., Hong, S.H. and Cho, K.S. (2014). Effects of water temperature and backwashing on bacterial population and community in a biological activated carbon process at a water treatment plant. Applied microbiology and biotechnology. 98: 1417-1427.

Kirchman, D.L. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. FEMS Microbiology Ecology. 39: 91-100.

Kim, T.G., Yun, J., Hong, S.-H. and Cho, K.-S. (2014). Effects of water temperature and backwashing on bacterial population and community in a biological activated carbon process at a water treatment plant. Applied Microbiology and Biotechnology. 98: 1417-1427.

Kirmeyer, G.J., Odell, L.H., Jacagelo, J., Wilczak, A. and Wolfe, R.L. (1995). Nitrification occurrence and control in chloraminated water systems. Denver CO: AWWA Research Foundation and America Water Works Association.

Koch, A.L. (2001). Oligotrophs versus copiotrophs. Bioessays. 23: 657-661.

Kovaleva, J., Degener, J.E. and Van der Mel, H.C. (2014). Methylobacterium and Its Role in Health Care-Associated Infection. Journal of Clinical Microbiology 52: 5.

Kowalchuk, G.A. and Stephen, J.R. (2001). Ammonia-oxidising bacteria: A model for molecular microbial ecology. Annual Reviews of Microbiology. 55: 485-529.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analysing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. 79: 5112-5120.

Kwon, S., Moon, E., Kim, T., Hong, S. and Park, H. (2011). Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. Microbes and Environments. 26(2): 149-155.

Lapara, T.M., Wilkinson, K.H., Strait, J.M., Hozalski, R.M., Sadowksy, M.J. and Hamilton, M.J. (2015). The bacterial communities of full-scale biologically active, granular activated carbon filters are stable and diverse and potentially contain novel ammonia-oxidizing microorganisms. Applied and environmental microbiology. 81: 6864-6872

Lautenschlager, K., Hwang, C., Liu, W.-T., Boon, N., Köster, O., Vrouwenvelder, H., Egli, T. and Hammes, F. (2013). A microbiology-based multiparametric approach towards assessing biological stability in drinking water distribution networks. Water Research. 47: 3015-3025.

Lautenschlager, K., Hwang, C., Ling, F., Lui, W.T., Boon, N., Köster, O., Egli, T. and Hammes, F. (2014). Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment. Water Research. 62: 40-52.

LeChevallier, M.W., Babcock, T W. and Lee, R.G. (1987). Examination and characterisation of distribution system biofilms. Applied and Environmental Microbiology. 53: 2714-2724.

LeChevallier, M.W., Cawthon, C.D. and Lee, R.G. (1988). Inactivation of biofilm bacteria. Applied and Environmental Microbiology 54: 2492-2499.

LeChevallier, M.W., Lowry, C.D. and Lee, R.G. (1990). Disinfecting biofilms in a model distribution system. Journal AWWA. 82(7): 87-99.

LeChevallier, M.W., Norton, C., Camper, A.K., Morin, P., Ellis, B., Jones, W., Rompe, A., Prevost, M., Coallier, J., Servais, P., Holt, D., Delanoue, A. and Colbourne, J. (1998). Microbial impact of biological filtration. American Water Works Association Research Foundation, Denver, Colo.

Lee, W.S., Yu, F.L. and Fang, W. (2014). Streptococcus parasanguinis coinfection with Escherichia coli Bacteremia in a patient with Complicated Urinary Tract Infection. Journal of Experimental and Clinical Medicine. 6: 2.

Lehtola, M.J., Miettinen, I.T., Vartiainen, T. and Martikainen, P.J. (2002). Changes in content of microbially available phosphorus, assimilable organic carbon and microbial growth potential during drinking water treatment processes. Water Research. 36: 3681-3690.

Legnani, P., Leoni, E., Soppelsa, F. and Burigo, R. (1998). The occurrence of Aeromonas species in drinking water supplies of an area of the Dolomite Mountains, Italy. Society for applied microbiology. 85: 6.

Lehtola, M.J., Miettinen, I.T., Keinanen, M.M., Kekki, T.K., Laine, O., Hirvonen, A., Vartiainen, T. and Martikainen, P.J. (2004). Microbiology, chemistry and biofilm development in a pilot drinking water distribution system with copper and plastic pipes. Water Research. 38: 3769-3779.

Lehtola, M.J., Laxander, M., Miettinen, I.T., Hirvonen, A., Vartiainen, T. and Martikainen, P.J. (2006). The effects of changing water flow velocity on the formation of biofilms and water quality in pilot distribution system consisting of copper or polyethylene. Water Research. 40: 2151-2160.

Lemarchand, C., Jardillier, L., Carrias, J. F., Richardot, M., Debroas, D., Sime-Ngando, T. and Amblard, C. (2006). Community composition and activity of prokaryotes associated to detrital particles in two contrasting lake ecosystems. FEMS microbiology ecology. 57: 442-451.

Levine, A.D., Tchobanoglous, G. and Asano, T. (1991). Size distributions of particulate contaminants in wastewater and their impact on treatability. Water Research. 25: 911-922.

Li, X., Upadhyaya, G., Yuen, W., Brown, J., Morgenroth, E. and Raskin, L. (2010). Changes in microbial community structure and function of drinking water treatment bioreactors upon phosphorus addition. Applied and Environmental Microbiology. doi:10.1128/AEM.01232-10.

Li, C., Ling, F., Zhang, M., Liu, W. T., Li, Y. and Liu, W. (2017). Characterization of bacterial community dynamics in a full-scale drinking water treatment plant. Journal of Environmental Sciences. 51: 21-30.

Liao, X., Chen, C., Chang, C.-H., Wang, Z., Zhang, X. and Xie, S. (2012). Heterogeneity of microbial community structures inside the up-flow biological activated carbon (BAC) filters for the treatment of drinking water. Biotechnology and bioprocess engineering. 17: 881-886.

Liao, X., Chen, C., Wang, Z., Wan, R., Chang, C.H., Zhang, X. and Xie, S. (2013). Pyrosequencing analysis of bacterial communities in drinking water biofilters receiving influents of different types. Process Biochemistry. 48: 703-707.

Liao, X., Chen, C., Wang, Z., Chang, C.H., Zhang, X. and Xie, S. (2015a). Bacterial community change through drinking water treatment processes. International Journal of Environmental Science and Technology. 12: 1867-1874.

Liao, X., Chen, C., Zhang, J., Dai, Y., Zhang, X. and Xie, S. (2015b). Operational performance, biomass and microbial community structure: impacts of backwashing on drinking water biofilter. Environmental Science and Pollution Research. 22: 546-554.

Lin, W., Yu, Z., Zhang, H. and Thompson, I.P. (2014). Diversity and dynamics of microbial communities at each step of treatment plant for potable water generation. Water Research. 52: 218-230.

Little, A.E., Robinson, C.J., Peterson, S.B., Raffa, K.F. and Handelsman, J. (2012). Rules of engagement: Interspecies interactions that regulate microbial communities. Annual Review of Microbiology. 62: 375-401.

Liu, B., Gu, L., Yu, X., Yu, G., Zhang, H. and Xu, J. (2012a). Dissolved organic nitrogen (DON) profile during backwashing cycle of drinking water biofiltration. Science of the Total Environment. 414: 508-514.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012b). Comparison of next-generation sequencing systems. BioMed Research International. doi:10.1155/2012/251364.

Lui, G., Verberk, J.Q.J.C. and Van Dijk, J.C. (2013a). Bacteriology of drinking water distribution systems: an integral and multidimensional review. Applied and Environmental Microbiology. 97: 9265-9276.

Liu, G., Lut, M.C., Verberk, J.Q.J.C. and Van Dijk, J.C. (2013b). A comparison of additional treatment processes to limit particle accumulation and microbial growth during drinking water distribution. Water Research. 47(8): 2719-2728.

Liu, G., Bakker, G.L., Vreeberg, J.H.G., Verberk, J.Q.J.C., Medama, G.J., Lui, M.C. and Van Dijk, J.C. (2014). Pyrosequencing reveals bacterial communities in unchlorinated drinking water distribution system: An integral study of bulk water, suspended solids, loose deposits and pipe wall biofilm. Environmental Science and Technology. 48: 5467-5476.

Liu, J., He, X.X., Lin, X.R., Chen, W.C., Zhou, Q.X., Shu, W.S. and Huang, L.N. (2015). Ecological effects of combined pollution associated with e-waste recycling on the composition and diversity of soil microbial communities. Environmental Science and Technology. 49: 6438-6447.

Llirós, M., Inceoğlu, Ö., García-Armisen, T., Anzil, A., Leporcq, B., Pigneur, L.-M., Viroux, L., Darchambeau, F., Descy, J.-P. and Servais, P. (2014). Bacterial community composition in three freshwater reservoirs of different alkalinity and trophic status. PloS one. 9: e116145.

Lopes, A.C.A., Martins, L.M., Gatti, M.S.V., Falavina dos Reis, C.M., Hofer, E. and Yano, T. (2015). Diarrhea Outbreak in Pernambuco, Brazil, Associated with a Heat-Stable Cytotoxic Enterotoxin Produced by Aeromonas Caviae. Rev Inst Med Trop Sao Paulo. 57: 4.

Lu, J., Struewing, I., Vereen, E., Kirby, A.E., Levy, K., Moe, C. and Ashbolt, N. (2016). Molecular Detection of Legionella spp. and their associations with Mycobacterium spp., Pseudomonas aeruginosa and amoeba hosts in a drinking water distribution system. Society for applied microbiology. 12.

Lui, G., Ling, F.Q., Mark, E.J., Zhang, X.D., Knezev, A., Verberk Meer, W.G.J., Madema, G.J., Luie, W.T. and Dijk, J.C. (2016). Comparison of particle associated bacteria from drinking water treatment plant and distribution reservoirs with different water sources. Scientific Reports. 6: 20-36

Luo, C., Tsementzi, D., Kyrpides, N., Read, T. and Konstantinidis, K.T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLOS One. 7: e30087.

Mains, C. (2008). Biofilm control in distribution systems. Tech Brief. 8: 1-4.

Magic-Knezev, A. and Van Der Kooij, D. (2004). Optimisation and significance of ATP analysis for measuring active biomass in granular activated carbon filters used in water treatment. Water Research. 38: 3971-3979.

Mara, D. and Horan, N. (2003). The Handbook of Water and Wastewater Microbiology. ACADEMIC PRESS, An imprint of Elsevier, School of Civil Engineering, University of Leeds, UK.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. Annual Review Genomics and Human Genetics. 9: 387-402.

Margulis, L. and Schwarts. (1988). Five kingdoms. Phyla on Earth 2: 2-8.

Martinez-Garcia, M., Swan, B.K., Poulton, N.J., Gomez, M.L., Masland, D., Sieracki, M.E. and Stepanauskas, R. (2012). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. The ISME Journal. 6: 113-123.

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.

Matilainen, A., Vepsäläinen, M. and Sillanpää. (2010). Natural organic matter removal by coagulation during drinking water treatment: A review. Advances in Colloid and Interface Science. 159: 189-197.

Mccarty, S.C. and Atlas, R.M. (1993). Effect of amplicon size on PCR detection of bacteria exposed to chlorine. Genome Research. 3: 181-185.

Menezes, F., Amal, R. and Luketina, D. (1996). Removal of particles using coagulation and flocculation in a dynamic separator. Powder Technology. 88: 27-31.

Metsämuuronen, S., Sillanpää, M., Bhatnagar, A. and Mänttäri, M. (2014). Natural organic matter removal from drinking water by membrane technology. Separation and Purification Reviews. 43: 1-61.

Metzker, M.L. (2005). Emerging technologies in DNA sequencing. Genome Research. 15: 1767-1776.

Metzker, M.L. (2010). Sequencing technologies—the next generation. Nature Reviews Genetics. 11: 31-46.

Mi, Z., Dai, Y., Xie, S., Chen, C. and Zhang, X. (2015). Impact of disinfection on drinking water biofilm bacterial community. Journal of Environmental Sciences. 37: 200-205.

Miller, H.C., Wylie, J., Dejean, G., Kaksonen, A.H., Sutton, D., Braun, K. and Puzon, G. J. (2015). Reduced efficiency of chlorine disinfection of Naegleria fowleri in a drinking water distribution biofilm. Environmental Science and Technology. 49: 11125-11131.

Mohammed, J.P. and Mattner, J. (2009). Autoimmune disease triggered by infection with alphaproteobacteria. Expert Rev Clin Immunol. 5: 18.

Mohammed, J.P., Fusakio, M.E., Rainbow, D.B., *et al*. (2011). Identification of Cd101 as a susceptibility gene for Novosphingobium aromaticivorans-induced liver autoimmunity. Journal Immunology. 187: 32.

Moll, D.M. and Summers, R.S. (1999). Assessment of drinking water filter microbial communities using taxonomic and metabolic profiles. Water Science and Technology. 39: 83-89.

Momba, M.N.B., Kfir, R., Venter, S.N. and Cloete, T.E. (2000). An overview of biofilm formation in the distribution systems and its impact on the deterioration of water quality. Water SA. 26(1): 59-66.

Morita, R.Y. (1997). Bacteria in Oligotrophic Environments. Starvation-survival Lifestyle. Chapman and Hall, New York. 529 pp.

Muyzer, G., De Waal, E.C. and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. Applied and Environmental Microbiology. 59: 695-700.

Ndiongue, S., Huck, P.M. and Slawson, R.M. (2005). Effects of temperature and biodegradable organic matter on control of biofilms by free chlorine in a model drinking water distribution system. Water Research. 39: 953-964.

Neden, D.G., Jones, R.J., Smith, J.R., Kirmeyer, G.J. and Foust, G.W. (1992). Comparing chlorination and chloramination for controlling bacterial regrowth. Journal AWWA. 84(7): 80-88.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'neill, S.P., Bilinski, T.M., Stanish, L.F., Knelman, J.E., Darcy, J.L., Lynch, R.C. and Wickey, P. (2013). Patterns and processes of microbial community assembly. Microbiology and Molecular Biology Reviews. 77: 342-356.

Nescerecka, A., Rubulis, J., Vital, M., Juhna, T. and Hammes, F. (2014). Biological instability in a chlorinated drinking water distribution network. PloS One. 9: e96354.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D. and Bertilsson, S. (2011). A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews. 75: 36.

Nicol, G.W. and Schleper, C. (2006). Ammonia-oxidising Crenachaeota: important players in the nitrogen cycle? Trends in Microbiology. 12(5): 207-212.

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

Nocker, A., Sossa-Fernandez, P., Burr, M.D. and Camper, A.K. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. Applied and environmental microbiology. 73: 5111-5117.

Nold, S.C. and Zwart, G. (1998). Patterns and governing forces in aquatic microbial communities. Aquatic Ecology. 32: 17-35.

Norton, C.D. and LeChevallier, M.W. (1997). Chloramination: its effect on distribution system water quality. Journal AWWA. 89(7): 66-77.

Norton, C.D. and LeChevallier, M.W. (2000). A pilot study of bacteriological population changes through potable water treatment and distribution. Applied and Environmental Microbiology. 66(1): 268-276.

Ollos, P.J. (1998). Effects of drinking water biodegradability and disinfectant residual on bacterial growth. Doctoral Thesis, University of Waterloo, Ontario.

Ollos, P.J., Huck, P.M. and Slawson, R.M. (2003). Factors affecting biofilm accumulation in model distribution systems. Journal AWWA. 95(1): 87-97.

O'melia, C.R., Hahn, M.W. and Chen, C.-T. (1997). Some effects of particle size in separation processes involving colloids. Water Science and Technology. 36: 119-126.

Organization WH. (2011). Microbial Fact Sheet. Guidelines for drinking-water quality, p. 84.

Organisation WH. (2014). Water Safety in Distribution Systems. WHO Document Production Services, Geneva, Switzerland.

Padgett, K.A., Selmi, C., Kenny, T.P., Leung, P.S.C., Balkwill, D.L., Ansari, A.A., Coppel, R.L. and Gershwin, E.M. (2005). Phylogenetic and immunological definition of four lipoylated proteins from Novosphingobium aromaticivorans, implications for primary biliary cirrhosis. Journal of Autoimmunity. 24: 11.

Pagano, T., Bida, M. and Kenny, J.E. (2014). Trends in levels of allochthonous dissolved organic carbon in natural water: A review of potential mechanisms under a changing climate. Water. 6: 2862-2897.

Pappu, N. and Patil, S.K. (2012). Water Purification Plant. Journal of Instrumentation and Control Engineering Department. 1: 36.

Percival, S.L., Knapp, J.S., Wales, D.S. and Edyvean, R.G.J. (1999). The effect of turbulent flow and surface roughness on biofilm formation in drinking water. Journal of Industrial Microbiology and Biotechnology. 22: 152-159.

Pernthaler, J., Zöllner, E., Warnecke, F. and Jürgens, K. (2004). Bloom of filamentous bacteria in a mesotrophic lake: identity and potential controlling mechanism. Applied and Environmental Microbiology. 70: 6272-6281.

Pintar, K.D.M. and Slawson, R.M. (2003). Effect of temperature and disinfection strategies on ammonia-oxidising bacteria in a bench-scale drinking water distribution system. Water Research. 37: 1805-1817.

Pindi, P.K., Yadav, P.R. and Shanker, A.S. (2013). Identification of opportunistic pathogenic bacteria in drinking water samples of different rural health centers and their clinical impacts on humans. BioMed Research International. 10.

Pinto, A.J., Xi, C. and Raskin, L. (2012a). Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environmental Science and Technology. 46: 8851-8859.

Pinto, A. and Raskin, L. (2012b). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLOS One. 7(8): e43093.

Pinto, A., Schroeder, J., Lunn, M., Sloan, W. and Raskin, L. (2014). Spatial-temporal survey and occupancy-abundance modelling to predict bacterial community dynamics in the drinking water microbiome. mBIO ASM. 5(3): e01135-14.

Pinto, A., Marcus, D.N., Ijaz, U.Z., Bautista-de los Santos, Q.M., Dick, G.J and Raskin, L. (2015). Metagenomic evidence for the presence of comammox Nitrospira-like bacteria in drinking water system. mSphere. 1(1): e00054-15.

Pinot, A., Marcus, D.N., Ijaz, U.Z., Bautista-de los Santos, M., Dick, G.J. and Raskin, L. (2016) Metagenomic evidence for the presence of commamox Nitrospira-like bacteria in a drinking water system. mSphere. 1(1): e00054-15. doi: 10.1128/mSphere.00054-15.

Prest, E.I., Hammes, F., Van Loosdrecht, M.C.M. and Vrouwenvelder, J.S. (2016). Biological stability of drinking water: controlling factors, methods and challenges. Frontiers in Microbiology. 7(45): doi: 10.3389/fmicb.2016.00045.

Prest, E.I., Weissbrodt, D.G., Hammes, F., Loosdrecht, M. C. M. and Vrouwenvelder, J.S. (2016). Long-Term Bacterial Dynamics in a Full-Scale Drinking Water Distribution System. PLoS ONE. 11: 20.

Proctor, C.R. and Hammes, F. (2015). Drinking water microbiology—from measurement to management. Current opinion in Biotechnology. 33: 87-94.

Redford, A.J. and Fierer, N. (2009). Bacterial succession on the leaf surface: a novel system for studying successional dynamics. Microbial ecology. 58: 189-198.

Regan, J.M., Harrington, G.W. and Noguera, D.R. (2002). Ammonia- and Nitrite-Oxidizing Bacterial Communities in a Pilot-Scale Chloraminated Drinking Water Distribution System. Applied and Environmental Microbiology. 68(1): 73-81.

Revetta, R.P., Gomez-Alvarez, V., Gerke, T.L., Curioso, C., Domingo, J.W.S. and Ashbolt, N.J. (2013). Establishment and early succession of bacterial communities in monochloramine-treated drinking water biofilms. FEMS Microbiology Ecology. 11.

Rice, E.W., Reasoner, D.J., Johnson, C.H.A. and DeMaria, L.A. (2000). Monitoring for Methylobacteria in Water Systems. Journal of Clinical Microbiology. 38: 2.

Richardson, A.E. and Simpson, R.J. (2011). Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiology. 156: 989-996.

Roeselers, G., Coolen, J., Van der Wielen, P.W.J.J., Jaspers, M.C., Atsma, A., deGraaf, B. and Schuren, F. (2015). Microbial biogeography of drinking water: patterns in phylogenetic diversity across space and time. Environmental Microbiology. 17(7): 2505-2514.

Rossman, L.A., Clark, R.M. and Grayman, W.M. (1994). Modeling chlorine residuals in drinking-water distribution systems. Journal of Environmental Engineering. 120(4): 803-820.

Roughgarden, J. (2009). Is there a general theory of community ecology? Biology and Philosophy. 24: 521-529.

Rutebemberwa, A., Stevens, M.J., Perez, M.J., Smith, L.P., Sanders, L., Cosgrove, G., Robertson, C.E., Tuder, R.M. and Harris, J.K. (2014). Novosphingobium and its potential role in chronic obstructive pulmonary diseases: Insights from microbiome msudies. PLOS ONE. 9: 13.

Schloss, P.D. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. PloS One 4: e8230.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H. and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 75: 7537-7541.

Schloss, P.D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Computational Biology. 6: e1000844.

Schloss, P.D. (2013). Secondary structure improves OTU assignments of 16S rRNA gene sequences. The ISME Journal. 7: 457.

Schloss, P.D., Gevers, D. and Westcott, S.L. (2011). Reducing the effects of PCR amplification and sequencing artefacts on 16S rRNA-based studies. PloS One. 6: e27310.

Schmeisser, C., Stockigt, C., Raasch, C., Wingender, J., Timmis, K.N., Wenderoth, D.F., Flemming, H.C., Liesegang, H., Schmitz, R.A., Jaeger, K.E. and Streit, W.R. (2003). Metagenome survey of biofilms in drinking water networks. Applied and Environmental Microbiology. 69(12): 7298-7309.

Schwartz, T., Kohnen, W., Jansen, B. and Obst, U. (2003). Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water bio¢lms. FEMS Microbiology Ecology. 43: 11.

Schwartz, T., Volkmann, H., Kirchen, S., Kohnen, W., Schon-Holz, K., Jansen, B. and Obst, U. (2006). Real-time PCR detection of Pseudomonas aeruginosa in clinical and municipal wastewater and genotyping of the ciprofloxacin-resistant isolates. FEMS Microbiology Ecology. 57: 10.

Scott, J. (2017). Role of Cyanobacteria in the ecosystem. Science. 1: 1-2

Shade, A., Caporaso, J.G., Handelsman, J., Knight, R. and Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. The ISME Journal. 7: 1493.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N. and Gilbert, J.A. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio. 5: e01371-01314.

Sharpton, T.J. (2014). An introduction to the analysis of shotgun metagenomic data. Frontiers in Plant Science. 5(209): 1-14.

Shokralla, S., Spall, J.L., Gibson, J.F. and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. Molecular Ecology. 21: 1794-1805.

Siqueira, V.M. and Lima, N. (2013). Biofilm formation by filamentous fungi recovered from a water system. Journal of Mycology. 1-9.

Soininen, J., Korhonen, J.J., Karhu, J. and Vetterli, A. (2011). Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. Limnology and Oceanography. 56: 508-520.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M. and Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidising archaea supports the phylum Thaumarchaeota. Trends in Microbiology. 18: 331-340.

Srinivasan, S., Harrington, G.W., Xagoraraki, I. and Goel, R. (2008). Factors affecting bulk to total bacteria ratio in drinking water distribution systems. Water Research. 42: 3393-33404.

Stackerbrand, E.R., Murray, G.E. and Truper, H.G. (1988). Proteobacteria classic novel name for the phylogenetic taxonomy that includes the Purple Bacteria and their relatives. International Journal of Systematic Bacteriology. 3: 321-325

Stahl, D.A. and de la Torre, J.R. (2012). Physiology and diversity of Ammonia-oxidising archaea. Annual Review of Microbiology. 66: 83-101.

Stevenson, D. (1994). The Specification of Filtering Materials for Rapid-Gravity Filtration. Water and Environment Journal, 8, 527-533.

Stewart, M. (2006). Acinetobacter. p. 7. American Water Works Association.

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

Szabó, K.É., Itor, P.O., Bertilsson, S., Tranvik, L. and Eiler, A. (2007). Importance of rare and abundant populations for the structure and functional potential of freshwater bacterial communities. Aquatic Microbial Ecology. 47: 1-10.

Tekerlekopoulou, A.G., Pavlou, S. and Vayenas, D.V. (2013). Removal of ammonium, iron and manganese from potable water in biofiltration units: a review. Journal of Chemical Technology and Biotechnology. 88: 751-773.

Thomas, J.M. and Ashbolt, N.J. (2011). Do Free-Living amoebae in treated drinking water systems present an emerging health risk? Environmental Science and Technology. 45(3): 860-869.

Thomas, T., Gilbert, J. and Meyer. (2012). Metagenomics – a guide from sampling to data analysis. Microbial Informatics Experimentation. 2(3): 2-12.

Tsai, Y. (2006). Interaction of chlorine concentration and shear stress on chlorine concentration, biofilm growth and particulate matter. Bioresorce Technology. 97: 1912-1919.

Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography. 33: 2-22.

Vaerewijck, M.J.M., Huys, G., Palomino, J.C., Swings, J.A. and Portaels, F. (2005). Mycobacteria in drinking water distribution systems: ecology and significance for human health. FEMS Microbiology reviews. 29: 24.

Van der Bruggen, B. and Vandecasteele, C. (2003). Removal of pollutants from surface water and groundwater by nanofiltration: overview of possible applications in the drinking water industry. Environmental Pollution. 122(3):435-445. https://doi.org/10.1016/S0269-7491(02)00308-1.

Van Nieuwenhuijzen, A., Van der Graaf, J., Kampschreur, M. and Mels, A. (2004). Particle related fractionation and characterisation of municipal wastewater. Water Science and Technology. 50: 125-132.

Van der Wende, E., Characklis, W.G. and Smith, D.B. (1989). Biofilms and bacterial drinking water quality. Water Research. 23(10): 1313-1322.

Van der Wielen, P.W.J.J., Voost, S. and Van der Kooij, D. (2009). Ammonia-oxidising bacteria and archaea in groundwater treatment and drinking water distribution systems. Applied and Environmental Microbiology. 75(14): 4687-4695.

Van der Wielen, P.W.J.J. and Van der Kooij, D. (2013). Nontuberculous Mycobacteria, Fungi, and Opportunistic Pathogens in Unchlorinated Drinking Water in the Netherlands. Applied and Environmental Microbiology (AEM). 79: 9.

Vasconcelos, J.L., Rossman, L.A., Grayman, W.M., Boulos, P.F. and Clark, R.M. (1997). Kinetics of chlorine decay. Journal AWWA. 89(7): 54-65.

Vaz-Moreira, I., Nunes, O.C. and Manaia, C.M. (2017). Ubiquitous and persistent Proteobacteria and other Gram-negative bacteria in drinking water. Science of the Total Environment. 9.

Vaz-Moreira, I., Egas, C., Nunes, O.C. and Manaia, C.M. (2012/2013) Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. FEMS Microbiology Ecology. 83: 14.

Vellend, M. (2010). Conceptual synthesis in community ecology. The Quarterly Review of Biology. 85: 183-206.

Velten, S., Hammes, F., Boller, M. and Egli, T. (2007). Rapid and direct estimation of active biomass on granular activated carbon through adenosine tri-phosphate (ATP) determination. Water Research. 41: 1973-1983.

Venables, W.N., Smith, D.M. and Team, R.D.C. (2004). An introduction to R. Network Theory Limited.

Vikesland, P.J., Ozekin, K. and Valentine, R.L. (2001). Monochloramine decay in model and distribution system waters. Water Research. 35(7): 1766-1776.

Vital, M., Dignum, M., Magic-Knezev, A., Ross, P., Rietveld, L. and 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 Research. 46: 4665-4676.

Volk, C.J. and LeChevallier, M.W. (1999). Impacts of the reduction of nutrient levels on bacterial water quality in distribution systems. Applied and Environmental Microbiology. 65(11): 4957-4966.

Von Gunten, U. (2003). Ozonation of drinking water: Part II. Disinfection and by-product formation in presence of bromide, iodide or chlorine. Water Research. 37: 1469-1487.

Vreeburg, I.J. and Boxall, J. (2007). Discolouration in potable water distribution systems: A review. Water Research. 41: 519-529.

Wang, J.Z., Summers, R. and Miltner, R.J. (1995). Biofiltration performance. I: Relationship to biomass. Journal-American Water Works Association. 87: 55-63.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology. 73: 5261-5267.

Wang, H., Edwards, M., Falkinham, J.O. and Pruden, A. (2012). Molecular survey of the occurrence of Legionella spp., Mycobacterium spp., Pseudomonas aeruginosa, and amoeba hosts in two chloraminated drinking water distribution systems. Applied and Environmental Microbiology. 78(17): 6285-6294.

Wang, H., Masters, S., Edwards, M.A, Falkinham, J.O. and Pruden, A. (2014). Effect of disinfectant, water age and pipe material on bacterial and eukaryotic community structure in drinking water biofilm. Environmental Science and Technology. 48: 1426-1435.

Wang, H., Bedard, E., Prevost, M., Camper, A.K., Hill, V.R. and Pruden, A. (2017). Methodological approaches for monitoring opportunistic pathogens in premise plumbing: A review. Elsevier. 117: 19.

White, C.P., Debry, R.W. and Lytle, D.A. (2012). Microbial survey of a full-scale, biologically active filter for treatment of drinking water. Applied and environmental microbiology. 78: 6390-6394.

WHO. (2004). Guidelines for Drinking-water Quality. Guidelines for Drinking-water Quality Vol. 1 p. 515. World Health Organization. Geneva.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis, Springer.

Widder, S., Allen, R.J., Pfeiffer, T., *et al*. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. The ISME Journal. 10: 2557.

Wilczak, A., Jacangelo, J.G., Marcinko, J.P., Odell, L.H., Kirmeyer, G.J. and Wolfe, R.L. (1996). Occurrence of nitrification in chloraminated distribution systems. Journal AWWA. 88(7): 74-84.

Wingender, J. and Flemming, H.C. (2004). Contamination potential of drinking water distribution network biofilms. Water Science and Technology. 49(11): 277-286.

Wingender, J. and Flemming, H.C. (2011). Biofilms in drinking water and their roles as reservoir for pathogens. International Journal of Hygiene and Environmental Health. 214: 417-423. doi:10.1016/j.ijheh.2011.05.009.

Wolfe, R.L., Lieu, N.I., Izaguirre, G. and Means, E.G. (1990). Ammonia-oxidising bacteria in a chloraminated distribution system: seasonal occurrence, distribution and disinfectant resistance. Applied and Environmental Microbiology. 56(2): 451-462.

Wu, Q.L., Zwart, G., Schauer, M., Kamst-van Agterveld, M.P. and Hahn, M.W. (2006). Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. Applied and Environmental Microbiology. 72: 5478-5485.

Wu, H., Zhang, J., Mi, Z., Xie, S., Chen, C. and Zhang, X. (2015). Biofilm bacterial communities in urban drinking water distribution systems transporting waters with different purification strategies. Environmental Biotechnology. 99: 1947-1955.

You, J., Das, A., Dolan, E. M. and Hu, Z. (2009). Ammonia-oxidising archaea involved in nitrogen removal. Water Research. 43: 1801-1809.

Youssef, N.H., Couger, M. and Elshahed, M.S. (2010). Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere. PLoS One. 5: e12414.

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.

Zarraonaindia, I., Smith, D.P. and Gilbert, J.A. (2013). Beyond the genome: community-level analysis of the microbial world. Biology and Philosophy. 28: 261-282.

Zavarzin, G., Stackebrandt, E. and Murray, R. (1991). A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. Canadian journal of microbiology. 37: 1-6.

Zeder, M., Peter, S., Shabarova, T. and Pernthaler, J. (2009). A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. Environmental microbiology. 11: 2676-2686.

Zeng, D.N., Fan, Z.Y., Chi, L., Wang, X., Qu, W.D. and Quan, Z.X. (2013). Analysis of the bacterial communities associated with different drinking water treatment processes. World Journal of Microbiology and Biotechnology. 29: 1573-1584.

Zhang, Y., Love, N. and Edwards, M. (2009). Nitrification in drinking water systems. Critical Reviews in Environmental Science and Technology. 39: 153-208.

Zhang, M., Lui, W., Nie, X., Li, C., Gu, J. and Zhang, C. (2012). Molecular analysis of bacterial communities in biofilms of a drinking water clearwell. Microbes and Environments. 27(4): 443-448.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., Martiny, J.B., Sogin, M., Boetius, A. and Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PLoS One. 6: e24570.

CAPACITY BUILDING

Diversity and dynamics of the microbial population associated with drinking water distribution networks and their impact on drinking water quality
WRC Report No. 2469/1/18

Capacity Building: Student training

Post-doctoral fellow:
Karen Jordaan
2015-2017

PhD:
Sarah MacRae (Potgieter)
2014-
Title: Functional diversity of bacteria in a drinking water distribution network

MSc:
S Vosloo
2014-2017
Title: Spatial and temporal patterns associated with prokaryotic communities in a drinking water treatment plant
Conference presentation:
Vosloo, S., Pinto, A.J., Crous, M., Sigidu, M., Du Preez, H. and Venter S.N. 2017. Spatial and temporal variability of prokaryotic communities in rapid sand filters used for drinking water treatment. Poster presentation at the 8th International Young Water Professionals Conference, Cape Town, South Africa, 10-13 December 2017.

M Crous
2016-2018
Title: Microbial community dynamics in a large drinking water treatment and distribution system
Conference presentation:
Vosloo, S., Pinto, A.J., Crous, M., Sigidu, M., Du Preez, H. and Venter S.N. 2017. Spatial and temporal variability of prokaryotic communities in rapid sand filters used for drinking water treatment. Poster presentation at the 8th International Young Water Professionals Conference, Cape Town, South Africa, 10-13 December 2017.

K Moodley
2016-2018
Title: Characterization of microbial communities in drinking water sources and distribution systems
Conference presentations:
Moodley, K. and Venter, S.N. 2017. Diversity and dynamics of the bacterial community associated with a water distribution network receiving limited treated ground water. Poster presentation at the 8th International Young Water Professionals Conference, Cape Town, South Africa, 10-13 December 2017.

Moodley K. and Venter S.N. 2018. Diversity and dynamics of the bacterial community associated with different ground and surface waters used for drinking water production. Poster to be presented at the SASM conference, Muldersdrift, 4-7 April 2018.

Hon:
Minette Crous
2015
Title: The diversity and spatial homogeneity of microbial communities associated with rapid sand filters used for drinking water treatment

Kinosha Moodley
2015
Title: Changes in the microbial community of chloraminated water in a municipal distribution and reticulation system

Luveshnie Gounden
2017
Title: Diversity and identification of potential bacterial pathogens in a large-scale drinking water distribution network in South Africa

Conference presentation:
Gounden L. and Venter S.N. 2018. Diversity and identification of potential bacterial pathogens in a large-scale drinking water distribution network in South Africa. Poster to be presented at the SASM conference, Muldersdrift, 4-7 April 2018.