

APPROACHES TO MONITOR AND CHARACTERIZE THE BIOLOGICAL STABILITY OF DRINKING WATER DISTRIBUTION NETWORKS

Volume II: Characterising the Composition of Microbial Communities and Correlations with Baseline FCM and HPC Cell Concentrations

Report to the
Water Research Commission

by

K Moodley¹, L Gounden¹, A Scholtz², I Loots² and SN Venter¹

¹Biochemistry, Genetics and Microbiology, University of Pretoria

²Civil Engineering, University of Pretoria

WRC Report No. 2884/2/23

ISBN 978-0-6392-0523-6

July 2023



Obtainable from

Water Research Commission
Bloukrans Building, 2nd Floor
Lynnwood Bridge Office Park
4 Daventry Street
Lynnwood Manor
PRETORIA

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

This report forms part of a set of three reports. The other reports are:

Approaches to Monitor and Characterize the Biological Stability of Drinking Water Distribution Networks.

Volume I: *Establishing Correlations between Flow Cytometry Cell Concentrations, Heterotrophic Plate Counts and Water Quality Data* (WRC Report No. 2884/1/23), and

Approaches To Monitor And Characterize The Biological Stability Of Drinking Water Distribution Networks.

Volume III: *Strategy for the Use of Flow Cytometry for Drinking Water Monitoring in South Africa* (WRC Report No. 2884/3/23).

DISCLAIMER

This report has been reviewed by the Water Research Commission (WRC) and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

EXECUTIVE SUMMARY

BACKGROUND

Within drinking water distribution networks, community reservoirs have been identified as one of the areas where bacterial regrowth can take place. Stagnation within these reservoirs could occur due to long residence times, sub-optimal flow dynamics and intermittent water demand. Stagnation is often associated with disinfectant decay resulting in an increase in bacterial numbers. These conditions often occur during the warmer summer months or during periods of water restrictions when a rapid deterioration of the microbial quality could put entire communities at risk.

RATIONALE

The major difficulty when implementing a direct assessment approach, such as FCM, within the water distribution environment is that there are no clear guidelines as to what constitutes a significant or relevant change in the microbial community. FCM counts have been shown to vary between different systems (chlorinated and chloraminated) and deviations or abnormal changes could only be detected once a proper baseline for both ICC and TCC values had been established for each system. During the study we therefore endeavoured to address the following questions:

- What is the composition of the microbial community (based on 16S profiling) of these samples and how does it correspond to the bacteria isolated using the HPC approach?
- Community reservoirs which form an important and integral part of the distribution network can be described as storage reservoirs which keep the balance between supply and demand in a distribution system. As these reservoirs store water, the water ages and the quality of the water can become questionable. Understanding the flow regimes and microbial ecology of community reservoirs will assist in the development of guidelines for the operation of these systems. For this reason, we also focused on the following questions:
 - What impact does retention time (e.g. flow restrictions) in community reservoirs have on disinfectant residuals, FCM values as well as the community composition?
 - What are the contributions of autotrophic bacteria to biological instability in the distribution system (as measured in reservoirs) and what functional role do they play in this ecosystem?
 - What are the main functions associated with reservoir communities as determined using a metagenomic approach?

OBJECTIVES AND AIMS

The objective of the project was to provide the necessary foundation for the development of a strategy for the drinking water industry to incorporate FCM when monitoring and managing the biostability of drinking water during distribution as this is a more sensitive and rapid method compared to the HPC currently used. The portion of the study focused on the composition of microbial communities in distribution networks and community reservoirs.

The project aims were to:

1. Investigate the main biological functions associated with communities that deviate from the baseline FCM values.
2. Investigate the contribution of autotrophic bacteria to biological instability in distribution system (as measured in reservoirs) and establish the functional role of these bacteria in the ecosystem.
3. Investigate the impact of increased retention time in community reservoirs on disinfectant residuals, FCM values and community composition in order to assist the development of procedural guidelines for the management of these reservoirs.

APPROACH

The first part of the project was to investigate the value of flow cytometry as a process indicator when managing water distribution networks. Samples were collected from the a large distribution network at six different sampling locations on a bi-weekly basis over a period of 8 months. For reticulation samples (point of use), water was collected from different residential locations in Tshwane district. Six points were sampled on a bi-weekly basis over the same period of 8 months.

Heterotrophic plate counts were performed using yeast extract and Reasoner's 2 agar (R2A) following standard protocols. Flow cytometry concentrations were determined using SYBR Green I and propidium iodide stains to obtain total and intact cell concentrations. The pH and chlorine concentration of the samples were also determined. In addition, part of each sample was concentrated with membrane filtration, followed by DNA extraction and 16S profiling. Based on colony differences at least 3 colonies were picked from the R2A plates of each sample. These bacteria were identified based on partial 16 rRNA sequencing and phylogenetic analyses. Based on this dataset potential correlations between the different parameters were analysed.

For the second part of the project permission was obtained from two of the local municipalities in Gauteng to include one of their community reservoirs in this study. Reservoir A was situated in an area known to experience water quality problems. Water leaving the purification plant is typically of acceptable quality but by the time it reaches the consumer, the water quality may have deteriorated significantly. This reservoir was sampled over several months to determine the interplay between design and flow patterns of the reservoir on the microbial quality within the reservoir.

With the help of the Department of Civil Engineering (UP), measurements such as flow, temperature and water levels within the reservoir were collected as input to create a crude model of the hydrology within the reservoir using the Computer Fluid Dynamics (CFD) system. This information was used to identify possible stagnation zones in the reservoir. This data served as input for the location of sampling points within the reservoir. To ensure that enough biological material was collected, all the samples were concentrated with membrane filtration. This was followed by DNA extraction from each sample which was used for 16S profiling and metagenome studies.

Sampling was also done at an additional community reservoir (Reservoir B) to determine the effect of residence time on the microbial water community in the reservoir. This reservoir received treated drinking water from a large water treatment works. After the filling of the reservoir, the inlet to the reservoir was closed and not refilled until the reservoir dropped to a level of 35%. This was done to allow for the longest possible residence time in the system. Sampling was conducted over a period of a week. Samples were processed and analysed as indicated above.

RESULTS AND DICUSSION

1. Investigating the composition of microbial communities in distribution systems

The 16S profiling results showed that the bacterial diversity is high amongst all these samples. The bacterial communities were rather unique among samples and the abundance of specific species varied. Various parameters could be responsible for the differences in the bacterial diversity across all sampling points. When looking at each sampling location, one would expect that the bacterial community present would be fairly consistent but a temporal influence played a vital role in the variation in the observed diversity. None of the samples where the FCM count deviated from the norm had any specific group dominating the sample.

A large diversity of bacteria was recovered from the plate counts as the selected isolates belonged to 53 genera and 28 families of which the majority are well known to be present in drinking water. The family *Sphingomonadaceae* had the most representatives with most belonging to the genus *Sphingomonas*. Several

others grouped with isolates from the genera *Sphingopyxis* and *Sphingobium*, which are known to be important environmental bacteria. The *Methylobacteriaceae* was the second most common family with isolates belonging to the genera *Methylobacterium* and *Methylorubrum*. Another prominent family was the *Mycobacteriaceae*. Some of these isolates grouped closely with some of the opportunistic pathogenic *Mycobacterium* species. The last prominent family was the *Comamonadaceae*. The diversity of these families varied across the sampling locations and some families had more isolates from specific points. Together they represented 2/3 of the total number of isolates obtained during this study.

2. Investigating the contribution of autotrophic bacteria to biological instability in distribution system

An analysis of Reservoir A was conducted to investigate the contribution of autotrophic bacteria to biological instability in distribution system and establish the functional role of these bacteria in the ecosystem. The assessment of the design of Reservoir A showed that depending on the fill-draw cycle, regions of stagnation could be predicted. It is believed that a late fill-draw cycle could have a larger stagnation zone directly opposite the inlet on the other side of the reservoir. The microbial population based on 16S profiling looked very similar at point and different depths throughout the reservoir indicating a homogenous bacterial community was present in the reservoir.

Proteobacteria and *Bacteroidetes* represented the highest abundances with members belonging to *Flavobacterium*, *Polynucleobacter*, *Burkholderiaceae*, *Porichthyaceae*, *Nitrospira* and *Sediminibacterium*. All these bacteria are commonly seen in drinking water systems. Genera such as *Flavobacterium* and *Nitrospira* are known to include autotrophic members.

The metagenome study of Reservoir A also showed that the *Proteobacteria* and *Bacteroidetes* were the most abundant members of the community. Deep taxonomy classification proved to be demanding of the bins compiled, indicating that most of them represented potential novel species, even genera. The only bin that could be identified to species level was Metabat240 which represented *Mycobacterium arupense*. This species is known as an opportunistic pathogen.

The functional analysis of the metagenome data revealed that the community performed all the basic enzymatic functions but that some enzymes were not present in any samples. Regardless of the month sampling was performed and although the species composition changed, the microbial functionality remained constant, showing microbial stability and functional redundancy.

3. Impact of increased retention time in community reservoirs on disinfectant residuals, FCM values and community composition

The reservoir community seemed to be similar regardless of the time of day the samples were taken. This implied that the retention time (tested for up to 3 days) didn't influence the community composition. Dominant bacteria seen at all sample points included *Nitrosomonas*, *Phrarobacter*, *Sphingomonas* and *Sphaerotilus*. These bacteria are common in drinking water systems.

CONCLUSIONS

During this project we studied the use of various analytical tools to study and ultimately manage the biostability of drinking water in distribution and reticulation systems. The analytical methods included culturing (HPC), an approach which has been used by the industry over many years as well as more recent molecular tools such as 16S profiling and metagenome analyses. We also explored flow cytometry as an additional process indicator. This project clearly demonstrated that these newer technologies have developed to such a level that they can now easily be incorporated into microbial drinking water quality studies. The costs associated with the sequencing-based techniques is also decreasing to a level where it can be considered for more routine use.

From a research perspective the vast amount of information which is collected when applying these molecular approaches provides a detailed view of the microbial community and its members. When compared with samples taken at other time points or sampling points this data could be used investigate the interactions and dynamics within the distribution system or reservoir. Combined with other water quality parameters this information provides a better understanding of the microbial ecology of such systems.

RECOMMENDATIONS

Interpretation and integration of the various sets of information and how to apply it when managing large networks remain the main challenge. Implementation of these analyses for routine purposes within the industry should only be considered after a careful cost benefit analysis. The main cost associated with these analyses is not necessarily linked to the direct costs of the analyses or the required infrastructure but often lies with the human resources component. This type of data interpretation requires a highly skilled team of scientists with a detailed understanding of the system, its associated microbiology as well as bioinformatic analyses.

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 V Akinwekomi	Magalies Water
Prof. TG Barnard	University of Johannesburg
Prof. C Bezuidenhout	North-West University
Dr N Gomba	NIOH
Mr J Kutu	City of Tshwane
Mr W Le Roux	CSIR
Mr V Masindi	Magalies Water
Ms A Maurizi	City of Ekurhuleni
Mr S Nduli	Magalies Water
Ms MV Sigudu	Rand Water
Dr V Ntema	Umgeni Water
Mr M Vulindlu	City of Cape Town

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.

page was intentionally left blank

TABLE OF CONTENTS

CHAPTER 1: BACKGROUND	1
1.1 INTRODUCTION	1
1.2 PROFILING MICROBIAL COMMUNITIES IN WATER DISTRIBUTION SYSTEMS	2
1.3 PROJECT OBJECTIVES AND AIMS	3
1.3.1 SCOPE AND LIMITATIONS	3
CHAPTER 2: MICROBIAL COMMUNITY PROFILING AT SELECTED DISTRIBUTION AND RETICULATION POINTS IN A WATER DISTRIBUTION SYSTEM	4
2.1 INTRODUCTION	4
2.2 MATERIALS AND METHODS	4
2.2.1 Sampling	4
2.2.2 Methods for determining the relative abundances of OTUs making up the communities	6
2.2.3 Methods for determining bacterial diversity in drinking water samples	8
2.3 RELATIVE ABUNDANCES OF OTUS MAKING UP THE COMMUNITIES	10
2.3.1 Taxonomic relative abundance distribution	10
2.3.2 OTU diversity and distribution	29
2.3.3 Alpha diversity	33
2.3.4 Beta diversity	34
2.4 BACTERIAL DIVERSITY IN DRINKING WATER SAMPLES BASED ON ISOLATES OBTAINED FROM HETEROTROPHIC PLATE COUNT ANALYSES	36
2.5 SUMMARY	37
2.5.1 Relative abundances of OTUs making up bacterial communities in the distribution systems	37
2.5.2 Bacterial diversity in drinking water samples based on isolates obtained from heterotrophic plate counts	39
CHAPTER 3: IMPACT OF THE RESERVOIR DESIGN AND MANAGEMENT ON THE MICROBIAL COMMUNITY	41
3.1 INTRODUCTION	41
3.2 MATERIALS AND METHODS	44
3.2.1 Reservoir selection and design	44
3.2.2 Analysis of Reservoir A flow dynamic	45
3.2.3 Sample collection for microbial profiling and functional analyses	53
3.2.4 Methods for microbial profiling of samples from Reservoir A	54
3.2.5 Community comparisons	56
3.2.6 Flow cytometry analysis	56
3.3 RESULTS AND DISCUSSION	56
3.3.1 Assessment of current design and flow patterns of the Reservoir A	56
3.3.2 Profiling the microbial community in Reservoir A	63
3.3.3 Flow cytometry data for Reservoir A	70
3.3.4 Functional role of dominant bacteria in a community Reservoir A	72
3.4 SUMMARY	86

CHAPTER 4: THE EFFECT OF RESIDENCE TIME ON THE MICROBIAL WATER COMMUNITY IN RESERVOIRS	87
4.1 INTRODUCTION	87
4.2 METHODS	87
4.2.1 Reservoir description	87
4.2.2 Sample collection and analysis	88
4.3 RESULTS AND DISCUSSION	88
4.3.1 Community membership/composition for Reservoir B	88
4.3.2 Alpha diversity	88
4.3.3 Beta diversity	90
CHAPTER 5: CONCLUSIONS	93
REFERENCES	94
APPENDIX A: SUPPLEMENTARY DATA FOR CHAPTER 2	103
APPENDIX B: SUPPLEMENTARY DATA FOR CHAPTER 3	110

LIST OF FIGURES

Figure 2.1: Geographical layout of the reticulation sampling locations.	6
Figure 2.2: Relative abundance plot per phyla for sample location Ben-Res2-IN over the sampling period....	11
Figure 2.3: Relative abundance plot per phyla for sample location Brak-RS3-IN over the sampling period.	11
Figure 2.4: Relative abundance plot per phyla for sample location Ga-Luka over the sampling period.	12
Figure 2.5: Relative abundance plot per phyla for sample location MB2 over the sampling period.	12
Figure 2.6: Relative abundance plot per phyla for sample location P4PL over the sampling period.	13
Figure 2.7: Relative abundance plot per phyla for sample location Rust-P6 over the sampling period.	13
Figure 2.8: Relative abundance plot per phyla for sample location AgricBuild over the sampling period.	14
Figure 2.9: Relative abundance plot per phyla for sample location Groenkloof over the sampling period.	15
Figure 2.10: Relative abundance plot per phyla for sample location NS2 over the sampling period.	15
Figure 2.11: Relative abundance plot per phyla for sample location Silverton over the sampling period.	16
Figure 2.12: Relative abundance plot per phyla for sample location Valhalla over the sampling period.	16
Figure 2.13: Relative abundance plot per phyla for sample location Waverley over the sampling period.	17
Figure 2.14: Relative abundance plot for the classes in phylum Proteobacteria for sample location Ben-Res2-IN over the sampling period.	18
Figure 2.15: Relative abundance plot for the classes in phylum Proteobacteria for sample location Brak-RS3-IN over the sampling period.	18
Figure 2.16: Relative abundance plot for the classes in phylum Proteobacteria for sample location Ga-Luka over the sampling period.	18
Figure 2.17: Relative abundance plot for the classes in phylum Proteobacteria for sample location MB2 over the sampling period.	19
Figure 2.18: Relative abundance plot for the classes in phylum Proteobacteria for sample location P4PL over the sampling period.	19
Figure 2.19: Relative abundance plot for the classes in phylum Proteobacteria for sample location RP6 over the sampling period.	19
Figure 2.20: Relative abundance plot for the classes in phylum Proteobacteria for sample location AgricBuild over the sampling period.	20
Figure 2.21: Relative abundance plot for the classes in phylum Proteobacteria for sample location Groenkloof over the sampling period.	20
Figure 2.22: Relative abundance plot for the classes in phylum Proteobacteria for sample location NS2 over the sampling period.	21
Figure 2.23: Relative abundance plot for the classes in phylum Proteobacteria for sample location Silverton over the sampling period.	21
Figure 2.24: Relative abundance plot for the classes in phylum Proteobacteria for sample location Valhalla over the sampling period.	22

Figure 2.25: Relative abundance plot for the classes in phylum Proteobacteria for sample location Waverley over the sampling period.	22
Figure 2.26: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Ben-Res2-IN over the sampling period.	23
Figure 2.27: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Brak-RS3-IN over the sampling period.	23
Figure 2.28: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Ga-Luka over the sampling period.	24
Figure 2.29: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample MB2 over the sampling period.	24
Figure 2.30: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample P4PL over the sampling period.	25
Figure 2.31: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample Rust-P6 over the sampling period.	25
Figure 2.32: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location AgricBuild over the sampling period.	26
Figure 2.33: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Groenkloof over the sampling period.	26
Figure 2.34: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location NS2 over the sampling period.	27
Figure 2.35: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Silverton over the sampling period.	28
Figure 2.36: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Valhalla over the sampling period.	28
Figure 2.37: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Waverley over the sampling period.	29
Figure 2.38: Heatmap representing Site 1 samples constructed using the absolute sequence abundances (log transformed) of 87 OTUs across all sampling locations which were selected based on their abundance within a sample (i.e. relative abundance threshold $\geq 1\%$). The Heatmap boxes were colored from red-to-blue to represent higher and lower abundances.	30
Figure 2.39: Heatmap representing Site 2 samples constructed using the absolute sequence abundances (log transformed) of 87 OTUs across all sampling locations which were selected based on their abundance within a sample (i.e. relative abundance threshold $\geq 1\%$). The Heatmap boxes were colored from red-to-blue to represent higher and lower abundances.	31
Figure 2.40: Heatmap constructed using the absolute sequence abundances (log transformed) of 87 OTUs across all sampling locations which were selected based on their abundance within a sample (i.e. relative abundance threshold $\geq 1\%$). The heatmap boxes were coloured from red-to-blue to represent higher and lower abundances.	32
Figure 2.41: Alpha diversity represented through Observed species, Chao1, Shannon and Simpson values for all samples taken at the 12 sampling points.	34
Figure 2.42: MDS/PCoA plot on weighted-UniFrac distance representing the beta diversity of the community relationship between 12 sampling locations over an 8-month period.	35

Figure 2.43: MDS/PCoA plot on unweighted-UniFrac distance representing the beta diversity of the community relationship between 12 sampling locations over an 8-month period.....	35
Figure 3.1: Water transport from treatment plant to the community.....	44
Figure 3.2: Reservoir outlay as seen (a) inside and (b) from the top.....	45
Figure 3.3: Detailed diagram of (a) inlet and (b) front view of inlet.....	45
Figure 3.4: Methods for describing mixing efficiency in service reservoirs.....	46
Figure 3.5: Depths of the thermistor as installed in the reservoir.....	47
Figure 3.6: Thermistor locations within the reservoir.....	47
Figure 3.7: Photo of the CR300 data logger used during the study.....	48
Figure 3.8: Photo of the ultrasonic flowmeter used during this study.....	48
Figure 3.9: CAD model of the side view of the reservoir.....	49
Figure 3.10: CAD model of the top view of reservoir.....	50
Figure 3.11: Initial temperature conditions.....	52
Figure 3.12: Age passive scalar based on the model.....	53
Figure 3.13: Inlet passive scalar based on the model.....	53
Figure 3.14: Schematic diagram representing the sampling points at 1 meter and 3 meters in Reservoir A... ..	54
Figure 3.15: Water levels linked to the fill-draw cycles.....	57
Figure 3.16: Inlet and outlet velocities.....	58
Figure 3.17: Theoretical residence time for 1 January to 6 December 2020.....	59
Figure 3.18: Detention time for the period of 1 January to 6 December 2020.....	59
Figure 3.19: Mixing times as calculated for one fill-draw cycle.....	61
Figure 3.20: Mixing times as calculated for the early and late fill-draw cycle.....	61
Figure 3.21: Volumetric exchange for days with one fill-draw cycle.....	62
Figure 3.22: Volumetric exchange for days with two fill-draw cycles.....	62
Figure 3.23: Representation of the mean relative abundance (MRA) of bacterial phyla of all sample points with their respective repeats, from month to month.....	64
Figure 3.24: Representation of the mean relative abundances based on sample location (depths) over the 5 different months of sampling.....	65
Figure 3.25: Shannon distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. pH ranged from 7 to 9. Higher values show increased richness in diversity.....	66
Figure 3.26: Pielou distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. pH ranged from 7 to 9. Values closer to 100 show a higher evenness distribution.....	67

Figure 3.27: Observed distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November, and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. Observed values range from 50 to 600, showing richness in diversity.	68
Figure 3.28: Principal coordination plot, showing groupings based on sampling months and depths using a Bray Curtis matrix.	69
Figure 3.29: Principal coordination plot, showing groupings based on sampling months and depths using a Jaccard matrix.	69
Figure 3.30: Graph showing total, intact and damaged flow cytometry cell counts.	70
Figure 3.31: Representation of the abundance's levels of the 7 different sampling points over the different sampling months (Reservoir A). Heatmap shows most dominate sequences over the 7 sample points over the different months.	71
Figure 3.32: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points.....	72
Figure 3.33: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - March.	72
Figure 3.34: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - August.	72
Figure 3.35: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - September.....	73
Figure 3.36: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - October.....	73
Figure 3.37: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - November.....	73
Figure 3.38: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - December.....	74
Figure 3.39: Represents transcripts per million for each bin against their sample points.....	81
Figure 3.40: Represents relative transcripts per million abundances of metagenomic bins.	82
Figure 3.41: Representation of KEGG pathways associated with the 45 high quality bins obtained from the study (the redder the clocks the more complete the pathway, pathways missing from all bins are excluded).	83
Figure 3.42: Summary of pathways associated with each sample which are representative of the bacterial community in Reservoir A at the time of sampling (the redder the clocks the more complete the pathway, complete pathways found in all communities are excluded).	84
Figure 4.1: Representation of the mean relative abundance (MRA) of bacterial phyla of all sample points. Points include source inlet (SI), reservoir outlet (RO) and tower outlet (TO). Samples were taken early (E), midday(M) and late(L).....	89
Figure 4.2: Alpha diversity representation of the time of day and alpha parameters. Shannon (top left) observed (top right) and pielou (bottom left).	89
Figure 4.3: Alpha diversity representation of the sample points and alpha diversity parameters. Shannon (top left) observed (top right) and pielou (bottom left).	90
Figure 4.4: PcoA plot, showing Bray Curtis matrix against time of samples taken.	91

Figure 4.5: PcoA plot, showing Jaccard matrix against time of samples taken. 91

Figure 4.6: Representation of the abundance's levels of different sampling points over the different sampling days (Reservoir B). 92

LIST OF TABLES

Table 3.1: Reservoir characteristics for the study reservoir 44

Table 3.2: Temperature sensor information (Campbell Scientific, 2018) 48

Table 3.3: Boundary conditions used to guide flow paths and define regions 50

Table 3.4: Physics models used for simulation 51

Table 3.5: Change in fill-draw cycles during the study 60

Table 3.6: Fill-draw information of a day with two fill-draw cycles..... 60

Table 3.7: Temperature analysis and the potential link to stagnation 63

Table 3.8: Key for pooled samples 74

Table 3.9: Representation of all retained bins with completeness above 90% and contamination below 5% contamination, including taxonomic classification of bins from the SqueetaMeta, GTDB-Tk and MiGA databases. 75

Table 3.10: Unique functional abilities either present or absent in some of the samples 85

ACRONYMS & ABBREVIATIONS

AOC	Assimilable organic carbon
ATP	Adenosine triphosphate
DWDS	Drinking water distribution system
DWTP	Drinking water treatment plant
FCM	Flow cytometry
HPC	Heterotrophic plate counts
ICC	Intact cell counts
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PI	Propidium iodide
SANS	South African National Standard
SG	SYBR Green
SLMB	Schweizer Lebensmittelbuch (Swiss Food Book)
TCC	Total cell counts
WHO	World Health Organisation

CHAPTER 1: BACKGROUND

1.1 INTRODUCTION

Drinking water safety is a major concern for water utilities worldwide, and various measures are implemented to ensure no outbreaks or severe disease incidents result from poor water quality (WHO, 2011; Gillespie *et al.*, 2014, Ma *et al.*, 2015). Ensuring biological stability, defined as the stability of the microbial population within drinking water from the treatment plant to consumer's taps, where the microbial community's composition and concentration does not change drastically throughout the system over time (Hammes *et al.*, 2010; Lautenschlager *et al.*, 2013; Nescerecka *et al.*, 2014; Prest *et al.*, 2016), is one of the main objectives. Monitoring of the system is an important activity when addressing biological stability. This is typically done by detecting the heterotrophic bacterial counts. This data is used for risk assessment and planning to help react effectively to issues detected in the system (WHO, 2004; 2011).

Disinfection and maintaining a disinfectant residual to ensure the microbial quality from post-treatment to end-point is often essential. The typical treatment used in drinking water systems includes chlorination and chloramination. Chlorine is mostly used as primary disinfectant as it has a high initial reactivity but decays throughout the system. A secondary disinfectant such as chloramine is therefore used to overcome this problem. Chloramine forms part of the total chlorine residual, has a lower reactivity and better effective concentrations and are thus better maintained throughout the system (Lu *et al.*, 2013; Zhu *et al.*, 2014; Prest *et al.*, 2016).

Culture-dependent and independent methods used to monitor the water quality assist water utilities in their efforts to manage the microbial quality of water. Culture-dependent methods have been relied upon for many years but fall short in many ways such as the lengthy time between sample collection and analysis. Due to problems with the cultivability and viability of certain bacteria, it is not possible to quantify and provide a complete overview of the microbial community present by only using a cultivation approach (Douterelo *et al.*, 2014; Tallon *et al.*, 2005; Ashbolt *et al.*, 2001).

The heterotrophic plate count is widely used as an operational parameter in water that could indicate treatment, repair and installation problems (Amanidaz *et al.*, 2015). Plate counts are often done at different temperatures, low temperatures to indicate the general dynamics in the system and higher temperatures to reflect possible contamination and the presence of pathogens (Sartory, 2004; Francisque *et al.*, 2009). The nutritional composition of the different media used also affects the microorganisms which grow. Heterotrophic bacteria grow well in conditions with higher temperature and pH and the concentrations in drinking water were reportedly between 1 CFU/mL and 10000 CFU/mL (Amanidaz *et al.*, 2015; Prest *et al.*, 2016; Sartory, 2004; Francisque *et al.*, 2009).

Culture-independent methods overcome the shortcomings of traditional methods in that they are rapid, the entire microbial community can be quantified, and in some cases, the pathogens present can also be detected. Flow cytometry has gained prominence as it can be used for site-specific monitoring. Flow cytometry technology rapidly provides an indication of the total and intact cells and is also sensitive to cell concentrations (Hammes *et al.*, 2008). Flow cytometry is a fluorescence-based technology which captures the data of stained cells that pass through a laser beam. A narrow stream of sheath fluid and stained cells are carried pass multiple lasers and as light is scattered off these cells it is amplified by photomultipliers and detected with detectors (Gillespie, 2016; Douterelo *et al.*, 2014; Selliah *et al.*, 2019; Rockey *et al.*, 2019; Van Nevel *et al.*, 2017; Hammes *et al.*, 2010; 2012; Prest *et al.*, 2013). De Roy *et al.* (2011) outlined additional advantages of flow cytometry which include multi-parameter analysis, single and multiple cell detection, high-

speed processing, high accuracy and no need for DNA extractions and amplifications. The ability of this technology to be highly reproducible with error rates less than 5% makes it a promising alternative or an additional parameter to heterotrophic plate counts (Prest *et al.*, 2014) to provide additional information on drinking water in distribution systems.

This portion of the study was designed to address the following questions:

- What is the composition of the microbial community (based on 16S profiling) of these samples and how does it correspond to the bacteria isolated using the HPC approach?
- Community reservoirs which form an important and integral part of the distribution network can be described as storage reservoirs which keep the balance between supply and demand in a distribution system. As these reservoirs store water, the water ages and the quality of the water can become questionable. Understanding the flow regimes and microbial ecology of community reservoirs will assist in the development of guidelines for the operation of these systems. For this reason, we also focused on the following questions:
 - What impact does retention time (e.g. flow restrictions) in community reservoirs have on disinfectant residuals, FCM values as well as the community composition?
 - What are the contributions of autotrophic bacteria to biological instability in the distribution system (as measured in reservoirs) and what functional role do they play in this ecosystem?
 - What are the main functions associated with reservoir communities as determined using a metagenomic approach?

1.2 PROFILING MICROBIAL COMMUNITIES IN WATER DISTRIBUTION SYSTEMS

Microbial communities are highly diverse and numerous factors influence the occurrence of different bacteria in water distribution systems. Municipalities and drinking water treatment facilities make use of a range of processes to eliminate the bulk of the microbial population to ensure the supply of safe and high-quality drinking water free of pathogens and contaminants to consumers (Lu *et al.*, 2013; Berney *et al.*, 2007; Tallon *et al.*, 2005). Various culture dependent methods are used to monitor the microbial community and the drinking water quality as is outlined in the Volume I report. The culturable community represents a small fraction of the total population and culture-independent methods based on sequencing data are employed to gain insight into the whole microbial community present in these systems.

The use of next-generation sequencing has led to an advanced understanding of the microbial composition and function in diverse environments and the data generated requires sophisticated computational analyses to interpret (Jovel *et al.*, 2016). The use of 16S rRNA gene profiling to investigate the diversity of bacterial communities provides a detailed overview of the heterogeneous population that shapes the microbial community (Lu *et al.*, 2013). Next-generation sequencing was shown to target the majority of unseen and uncultivable microbes present in the distribution system (Van Nevel *et al.*, 2017). Our current understanding of the microbial diversity in water systems is therefore primarily based on 16S rRNA gene sequences as is a target region shared by all bacteria (Fuks *et al.*, 2018).

The diversity of microbial communities varies, and it is widely observed that only a few species are present at high abundances. In contrast many different species are present at a low abundance (Ferrenberg *et al.*, 2013). These rare species may sometimes increase in abundance if conditions are favoured and these species are shown to be an important indicator of environmental disturbances. In water distribution systems, Proteobacteria were shown to be the dominant phyla and *Oxalobacteraceae* and *Methylobacteriaceae* as the most abundant bacterial families (Zhang, 2012). Biofilms which form on the pipe walls are said to be dominated by very few taxa which vary in abundance at different sampling times and conditions (Kelly *et al.*, 2014; Props *et al.*, 2018). The relative abundances of operational taxonomic units (OTUs) representing a

community as well as species distribution are known to explain a community's structure (Zhang, 2012).

Microbial ecology is represented as either alpha or beta diversity to define and compare communities. Alpha diversity represents the diversity within a sampling point/location and evaluates the depth of sampling within a system or environment (Holinger *et al.*, 2014, Li *et al.*, 2016; Bautista-de los Santos *et al.*, 2016). In contrast, the beta diversity provides a measure of the structure of the community and is used to compare the diversity between sampling locations (Bautista-de los Santos *et al.*, 2016; Lu *et al.*, 2013; Jost, 2007). Alpha and beta diversity measures are defined in terms of the species abundances and species richness and different diversity indices and metrics can be employed (Jost, 2007).

1.3 PROJECT OBJECTIVES AND AIMS

The objective of the project was to provide the necessary foundation for the development of a strategy for the drinking water industry to incorporate FCM when monitoring and managing the biostability of drinking water during distribution as this is a more sensitive and rapid method compared to the HPC currently used. The project also focused on the impact of community reservoirs on the microbial quality of drinking water supplied to consumers. The project aims were to:

- 1 Create a baseline FCM (TCC and ICC values and fluorescent fingerprints), HPC (YEA and R2A) and 16S community profile databases for water samples from chloraminated distribution and reticulation networks.
- 2 Investigate the main biological functions associated with communities that deviate from the baseline FCM values.
- 3 Investigate the impact of increased retention time in community reservoirs on disinfectant residuals, FCM values and community composition in order to assist the develop procedural guidelines for the management of these reservoirs.

1.3.1 SCOPE AND LIMITATIONS

The main focus of the flow cytometry study involved distribution samples from a system operated by a large water utility as well as reticulation samples collected from different residential locations at the point of use. The study addressing the microbial dynamics of a reservoir focused on a community reservoir receiving water from a small-scale conventional treatment plant. The reservoir was selected as it is located in an area known to experience water quality issues. The impact of retention time on the microbial community of a reservoir targeted a different reservoir that formed part of a larger treatment and distribution network. Although all these systems were from the Gauteng 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: MICROBIAL COMMUNITY PROFILING AT SELECTED DISTRIBUTION AND RETICULATION POINTS IN A WATER DISTRIBUTION SYSTEM

2.1 INTRODUCTION

The microbial diversity within sampled datasets are distributed randomly due to the variation which may occur in the community. Through the measurement of the different types of diversity, insight is gained within and between sample sites regarding the species presence, species abundance, evenness and structure. Microbial communities can vary in composition which is influenced by treatment events and contamination through pollution (Korajkic *et al.*, 2015). In drinking water distribution systems, the structure of the system as well as various abiotic factors such as nutrients, pH and temperature contribute to shaping the microbial community; and the presence of biofilms which contain cyclic reactions influence variation in datasets (Lu *et al.*, 2013; Mara and Horan, 2003, Rubuli *et al.*, 2010).

Culture-dependent methods are used by most drinking water utilities to identify the presence and abundance of bacteria (Lu *et al.*, 2013). The utilities focus primarily on bacterial plate counts but the identities of the specific bacterial cultures are not often determined. The type of media, incubation temperature and time which is used for culturing these bacteria also play a role in the type of bacteria that grow. The effect of different heterotrophic plate count (HPC) methods on the culturable microbial community was investigated by Gensberger *et al.* (2015) and the study found that the abundance of particular microbes was affected by the method used. The study also highlighted that the use of culture-dependent methods may fail in the detection of rare taxa (Gensberger *et al.*, 2015). As the culturable community is only a small representation of the entire community (Berney *et al.*, 2007; Tallon *et al.*, 2005) present, culture-independent methods are also used to study the whole microbial community.

This study served two purposes;

- To investigate the relative abundances of OTUs making up the community at the 2 different sampling locations. The communities were put through alpha- and beta-diversity tests to investigate the relationships amongst the microorganisms. Heatmaps were constructed to investigate the shifts in the OTUs making up the microbial community part of treated drinking water samples collected from two study sites.
- To investigate the diversity in the culturable bacterial community, based on heterotrophic plate count isolates and investigate the dominant species at the sampling locations. To obtain the identity of the isolates, the 16S r RNA genes were amplified after DNA extraction and sequenced using the Sanger sequencing approach. Final identification was performed by means of phylogenetic comparisons.

2.2 MATERIALS AND METHODS

2.2.1 Sampling

Samples for this purpose were collected at six different sampling locations of a large water distribution network on a bi-weekly basis over a period of 8 months. For reticulation samples (point of use), water was collected from different residential locations in Tshwane district. Six points were sampled on a bi-weekly basis over the same period of 8 months.

2.2.1.1 *Study site 1*

Samples were collected from a large distribution network at six different sampling locations on a bi-weekly basis over a minimum period of 6 months. The six sites were randomly chosen based on an initial set of 20 samples collected at various locations along the distribution system. The samples included reservoirs with potentially high bacterial counts and points along the distribution network. The six sample points were: MB2 (production), Ga-Luka (GL), P4PL, Rust-P6 (RP6), Brak-RS3-IN (BR3) and Ben-Res2-IN (BR2). MB2 represented a production sample taken directly after the final treatment and was included as control and for comparative purposes. For all samples, the source water was a canal from the Vaal dam and treatment and purification consisted of seven stages namely: coagulation; flocculation; sedimentation; stabilization; filtration; disinfection and chloramination. Samples MB2 and P4-PL were collected from completely different sections of the distribution network compared to RP6, GL, BR2, and BR3 which belonged to the same section of the distribution network. Samples were collected in sterile 8 L Large Narrow Mouth Nalgene polycarbonate bottles (Thermo Scientific™, South Africa). The samples were collected from the end of July 2018 to March 2019.

2.2.1.2 *Study site 2*

For reticulation samples, water was collected from different residential locations at the point of use in the Pretoria area (Figure 2.1). Six points were sampled on a bi-weekly basis over a minimum period of 6 months. The six sites were randomly chosen from an initial set of 10 samples based on the results of FCM and plate concentrations to include both high and low concentrations. The six sample points were: Waverly (WAV), Silverton (SIL), Groenkloof (GK), Valhalla (VAL), Natural Sciences building 2 (NS2) and Agricultural building (AB) (both NS2 and AB were in Hatfield on the main campus of the University of Pretoria). Of the six samples, four were obtained from household taps and two were obtained from taps in buildings on the University of Pretoria Hatfield campus. These two, AgricBuild and NS2 were from taps linked to storage tanks on the building's rooftop. The samples obtained from household taps were obtained from locations within the range of 10-30 km apart. Samples were collected in sterile 8 L Large Narrow Mouth Nalgene polycarbonate bottles (Thermo Scientific™, South Africa).

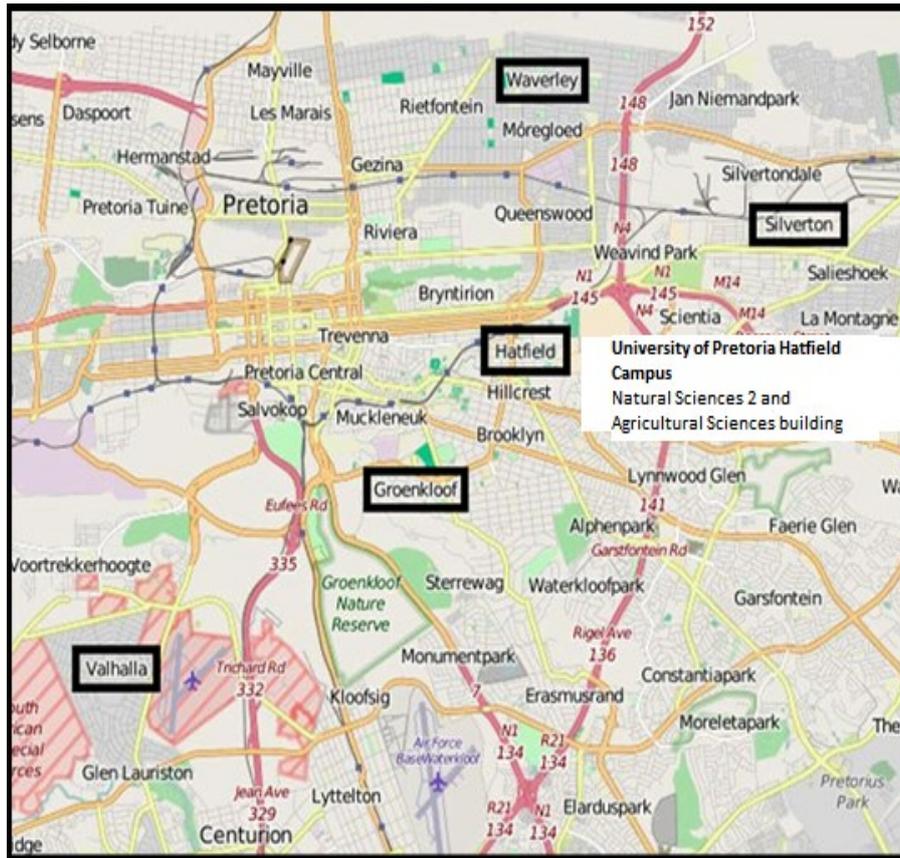


Figure 2.1: Geographical layout of the reticulation sampling locations.

2.2.2 Methods for determining the relative abundances of OTUs making up the communities

2.2.2.1 Phenol-chloroform genomic DNA Extraction

Genomic DNA was extracted from the twelve samples described in Section 2.2.1. The 8 L drinking water samples' microbial biomass was collected in Sterivex™ GP 0.22 µm polycarbonate filter units (Millipore) through filtration using the Gilson Miniplus 2 peristaltic pump. These filters were cut into multiple pieces using aseptic methods and then placed into 2 mL Lysing matrix Tubes E (MP Biomedical, South Africa) using tweezers. Genomic DNA was extracted using the phenol-chloroform method as described by Urakawa *et al.* (2010) with modifications incorporated by Feinstein *et al.* (2009) and Pinto *et al.* (2012). A volume of 300 µl of 2 x TENS buffer (100 mM Tris-HCl, 40 mM EDTA, 200 mM NaCl, 2% SDS) and 900 µL phenol: chloroform: isoamyl alcohol (25:24:1, pH 8) was added to the filter pieces in the 2 mL Lysing matrix Tube E. The heterogeneous mixture was then vigorously vortexed and put through three rounds of consecutive bead-beating in the Retsch Oscillating Mill MM 301 at the highest frequency (setting 6) for 40 sec. After the first round of bead-beating, the homogenous 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 new 2 ml Eppendorf tube. The original 2 ml Lysing Matrix Tube E tubes containing the pieces of filter was re-filled with 200 µL TENS buffer before the second round of bead-beating. The first two rounds of bead-beating were centrifuged for 10 min at 14 000 x g, the aqueous phase was transferred to the new 2 ml Eppendorf tube and 200 µL TENS buffer was added before the next bead-beating round. For the last round of bead-beating, the mixture was centrifuged for 10 min at 12 500 x g.

The aqueous phase accumulated after three rounds of bead-beating was approximately 600 µL. To collect the

aqueous phase containing the genomic DNA, 700 μ L phenol: chloroform: isoamyl alcohol (Sigma Aldrich, South Africa) was added to the 2 ml Eppendorf tube and thoroughly mixed by repeated inversions and then centrifuged using the Eppendorf Centrifuge 5804R™ at 14 000 x g for 5 min. The resulting aqueous phase (approximately 700 μ L) was then transferred to a new 2 ml Eppendorf tube and combined with (350 μ L) 7.5 M ammonium acetate (Sigma Aldrich, South Africa) and 600 μ L chloroform (Sigma Aldrich, South Africa). The tubes were mixed thoroughly through repeated inversions and thereafter centrifuged 14 000 x g for 5 min. The resulting aqueous phase was transferred to a new 2 ml Eppendorf tube and incubated at -80°C for 10 min after the addition of 600 μ L isopropanol (Sigma Aldrich, South Africa) and 1.5 μ L GlycoBlue™ co-precipitant (15 mg/ml) (Thermo Fischer Scientific, South Africa). To precipitate the nucleic acid, the incubated mixture was centrifuged for 30 min at 12 000 x g at 4°C using the Eppendorf Centrifuge 5804R™. After centrifugation, the supernatant was carefully removed without dislodging the blue pellet and 1 ml 80% ethanol was added to wash the pellet prior to centrifuging for another 30 min at 4°C using the Eppendorf Centrifuge 5804 R™. The supernatant was carefully removed and the pellet was air-dried for 5-10 min. The pellet was re-suspended in 50 μ L nuclease free water (Qiagen, South Africa) and stored at -20°C until further processing.

2.2.2.2 16S rRNA gene amplification

The genomic DNA small ribosomal ribonucleic acid (rRNA) encoding gene region was amplified using the methodology described in section 3.2.1.2. Successfully amplified samples were then checked for their DNA concentration and purity using the Nanodrop ND-1000™ Spectrophotometer (Thermo Scientific, South Africa) followed by Qubit. The nanodrop was used to establish the concentration of DNA in the sample and Qubit™ dsDNA BR assay kit (100 μ g/ μ L to 1000 ng/ μ L) was used with the Qubit® 2.0 fluorometer to quantify the DNA concentration more accurately. The Qubit kit consisted of Qubit™ dsDNA HS Reagent, Qubit™ dsDNA HS Buffer, Qubit™ dsDNA HS Standard #1 and Qubit™ dsDNA HS Standard #2. The DNA was quantified by diluting the reagent using the buffer and adding 1 μ L of sample which was incubated together for 2 min. The QF value was then read using the Qubit® 2.0 fluorometer, the QF value was used to calculate the concentration of the sample.

2.2.2.3 Next generation sequencing and data processing

The successfully amplified samples 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. Sequencing using a pair-ended sequencing approach as described by Kozich *et al.* (2013) was performed.

A total of 94 samples were successfully sequenced. The analysis of these samples was performed using The Divisive Amplicon Denoising Algorithm (DADA2) described by Callahan *et al.* (2016). The amplicon processing workflow included sequence filtering, dereplication, trimming, inferring sample composition, chimera identification and removal, merging paired-end reads, construction of a sequence table and assignment of taxonomic classifications. The original trimming and filtering of the reads followed the standard filtering parameters described for Illumina MiSeq 2x250 V4 region of the 16S rRNA gene (<https://benjjneb.github.io/dada2/tutorial.html>).

The reads with ambiguous bases were removed (maxN=0), the maximum number of errors expected was defined (maxEE=2) and the reads were truncated at a quality score less than or equal to truncQ (truncQ=2). Dereplication, where identical sequences are combined into “unique sequences” was then performed while the abundance of the number of reads corresponding to the unique sequences was maintained. The core sample inference algorithm was applied to the unique sequences and the forward and reverse reads were merged to obtain complete sequences (Callahan *et al.*, 2016). The merged reads were used to construct a sequence table, chimeras were removed after identification and taxonomic assignments were called using the

SILVA reference database at a species level (<https://benjjneb.github.io/dada2/training.html>) through DADA2 taxonomy script for assignment of taxa.

2.2.2.4 *Bacterial community analyses*

After taxonomic assignment with the DADA2 (Version 1.12.1) package, the sequences were processed further using packages installed in RStudio (Version 1.2.1335). The bacterial community relationships using alpha and beta diversity analyses were carried out. The Phyloseq package (Version 1.28.0) was used to analyse the microbial communities present from the sequences processed in DADA2. The workflow was followed and modified as required (https://vaulot.github.io/tutorials/Phyloseq_tutorial.html) and alpha and beta diversity were calculated as described by McMurdie and Holmes (2013). A Heatmap was drawn after running multiCoLA following the method described by Gobet *et al.* (2010) on the sequences and OTUs that had over an average of 50 sequences and more per OTU from the 1% abundance table output (Potgieter *et al.*, 2018). All plots were drawn using the ggplot package (Version 3.2.1). Beta diversity consisting of weighted and unweighted Unifrac analysis ordinated with an MDS/PCoA plot was run using the phyloseq package Fast Unifrac function. The relative abundance was calculated and stacked plots were drawn in Microsoft Excel 2013 per sample location at the phyla, class and family level.

2.2.3 **Methods for determining bacterial diversity in drinking water samples**

2.2.3.1 *HPC cultures*

Single colonies were picked based on the physical appearance on Reasoner's 2 agar (R2A) plates. At least three colonies based on differences in colour, texture and abundance on the initial plate were selected and re-streaked to obtain pure cultures. The isolated colonies were streaked on nutrient agar, checked for single colonies and were re-streaked to confirm that they are pure cultures. Colonies isolated from R2A were grown on R2A media and incubated at 22°C for 3-5 days to allow optimal growth of the isolates. Colonies isolated from yeast extract agar (YEA) were grown on nutrient agar and incubated at 35°C for a period of ± 44 hours.

2.2.3.2 *DNA extraction*

DNA was isolated from the pure cultures using either heat treatment (boiling) or a DNA extraction kit. For the DNA extraction using boiling, one to three colonies were selected from the plate and suspended into 50 µl of autoclaved distilled water in a 1.5 ml Eppendorf tube. The samples were then boiled in a heating block at 100°C for 10 min and thereafter centrifuged at 5000 rpm for 2 min using the Eppendorf Centrifuge 5804R™. DNA was taken from the supernatant for downstream analysis and stored at -20°C.

For isolates where the DNA was unsuccessfully extracted by boiling and those extracts that did not amplify with PCR, extraction was done using the Quick-gDNA™ Miniprep (ZYMO) kit. Colonies were scraped from the pure culture and suspended in 500 µl of genomic lysis buffer and vortexed using the Heidolph REAX 2000. The suspension was then allowed to stand for 5-10 minutes at room temperature. Thereafter the mixture was transferred to a Zymo-spin column in a collection tube and centrifuged at 10000g for 1 min, the flow-through was discarded and 200 µl of DNA pre-wash buffer was added to the spin column and centrifuged at 10000g for 1 min. The flow-through was discarded and 500 µl of gDNA wash buffer was added to the spin column and centrifuged at 10000 g for 1 min. The Zymo-spin column was transferred to a clean 1.5 ml Eppendorf tube and 30 µl of DNA elution buffer was added to the Zymo-spin column and incubated for 2-5 minutes at room temperature and thereafter centrifuged at maximum speed for 30 seconds to elute the DNA which was stored at -20°C until used.

2.2.3.3 16S rRNA gene amplification

The extracted genomic DNA was used to amplify the 16 S rRNA region using universal primers that specifically target this region. The forward and reverse primer based on Edwards *et al.* (1989) was used, namely 27F (5-AGA GTT TGA TCC TGG CTC AG-3') and 1492R (5'-GGT TAC CTT GTT ACG ACT-3'). The BIO-RAD T100™ Thermal Cycler was used to perform all PCRs. The PCR mixture with a final volume of 25 µl consisted of 1 x reaction buffer, 1.50 mM MgCl₂, 250 µM of each nucleotide (dATP, dCTP, dGTP, dTTP), 10 pmol of each primer (forward and reverse), 1.50 U Taq DNA polymerase (Super-therm Polymerase Jmr-801), 16.2 µl nuclease-free water (Qiagen, South Africa) and 1 µl genomic DNA. The PCR cycling conditions followed were an initial denaturation step at 92°C for 10 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, extension at 72°C for 1 min, and a final extension at 72°C for 5 min, at the end of the 30 cycles. Afterwards the reaction was kept at 4°C. The PCR product containing the amplicons were checked against a positive control consisting of *Escherichia coli* DNA on an agarose gel. For electrophoresis, a 1% agarose gel was made using 1X TAE buffer and distilled water. The gel was stained using SYBR Green DNA gel stain before pouring into the gel tank to set. For loading 2 µl of loading dye (Thermo Scientific™) was combined with 3 µl of the PCR product. The gel was run at 80 V for 20-25 min in the PowerPac™ Basic gel tank (Bio-Rad) and viewed on the Gel Doc™ EZ Imager (Bio-Rad) and data was captured using the Image Lab program (Version 3.0 build 1.1). The PCR was repeated for samples with failed amplification. For these samples, the DNA was diluted in a 1:10 ratio with nuclease-free water (Qiagen, South Africa) to remove potential inhibitors (e.g. ionic detergents, phenolic compounds, and ethanol).

2.2.3.4 Sanger Sequencing

The successfully amplified DNA was cleaned up before downstream analysis for sequencing PCR and Sanger sequencing. PCR clean-up was done on successfully amplified samples with Exonuclease I (20 U/µl) (Thermo Scientific™) and FastAP Thermosensitive Alkaline Phosphatase (1 U/µl) (Thermo Scientific™). The Exonuclease I (20 U/µl) is used to degrade any single-stranded DNA remaining as well as excess primers remaining before sequencing. The FastAP is used to remove any phosphate groups from the 5'- and 3' ends of DNA. The PCR product of approximately 20 µl was combined with 0.5 µl Exonuclease I and 2 µl FastAP and then incubated at 37°C for 15 min and at 85°C for 15 min.

After clean-up, a sequencing PCR was set up. The 10 pmol forward primer stock (27F (5-AGA GTT TGA TCC TGG CTC AG-3')) of volume 0.3 µl was used in combination with 2 µl 5X Sequencing buffer, 1 µl Big dye 3.1, nuclease-free water of volume 4.7 µl and 4 µl of the clean PCR product. The reaction cycles consisted of 25 cycles of initial denaturation at 96°C for 5 sec, denaturation at 96°C for 10 sec, annealing at 55°C for 5 sec and extension at 60°C for 4 min, at the end of the 25 cycles, the reaction was kept at 4°C.

After the sequencing PCR, the DNA was precipitated using sodium acetate/ethanol precipitation. The precipitation is used to concentrate the DNA present and remove any possible contaminants. The use of ethanol together with sodium acetate is to introduce ions to the mix and concentrate the DNA. For the precipitation, 16 µl of 100 % ethanol and 2 µl of sodium acetate pH 4.8 was added to 12 µl of sequencing product in 0.5 µl sequencing tubes. The homogenous mixture was then centrifuged at max speed in the Hermle Z 200™ centrifuge for 30 min. Two rounds of 5 min centrifuging with 150 µl of 70 % ice-cold ethanol were done and the supernatant was removed after each round, the pellet was washed. The tubes were then placed on a heating block for 3 min at 90°C with the caps open to evaporate any excess ethanol and thereafter stored at -20°C until sequenced.

The successful amplicons were sent to the University of Pretoria's DNA Sanger Sequencing Facility to be sequenced using the ABI3500xl Genetic Analyzer. The forward read of the 16S rRNA V4 variable region was

targeted with the 27F (5-AGA GTT TGA TCC TGG CTC AG-3') primer and a short run (< 750 bp) was done using Big dye 3.1 short.

2.2.3.5 Phylogenetic analyses

To identify the isolates, the sequences obtained from the Sanger Sequencing Facility were trimmed and checked in Bioedit Version 7.2.5. The isolate sequences' potential identities were obtained using NCBI Basic Local Alignment Search Tool (BLAST®). A standard nucleotide blast was conducted with parameters excluding models and uncultured/environmental sequences and limited to sequences from type material in the nucleotide database. Reference sequences of type strains closely related to the presumptive identity based on blast were obtained from the 16S rRNA LTPs132_SSU database and the List of prokaryotic names with standing in nomenclature (LPSN, <http://www.bacterio.net/>) was used to ensure that all the closely related validly published species are included. The isolate sequences obtained after Sanger sequencing were combined with the type strain sequences from the families which were associated with these sequences after BLAST identification. These sequences were combined using Bioedit and aligned online using MAFFT Version 7 (<https://mafft.cbrc.jp/alignment/server/>). Maximum likelihood phylogenetic trees were constructed using MEGA Version X using 500 bootstrap replicates.

2.3 RELATIVE ABUNDANCES OF OTUs MAKING UP THE COMMUNITIES

2.3.1 Taxonomic relative abundance distribution

From the 16S data generated, the relative abundance calculated at a phyla, class and family level are represented by Figures 2.2-2.37. At the phyla level (Figures 2.2-2.13), Proteobacteria was generally the most dominant phylum and was consistently seen across all the sampling locations but the specific phyla distribution often varied across sample locations. The dominant classes within Proteobacteria (Figures 2.14-2.25) were *Alpha*-Proteobacteria and *Gamma*-Proteobacteria. The families within these dominant classes and the bacterial diversity was investigated and are represented by Figures 2.26-2.37. The diversity of families found at each sampling location varied and the percentage of families observed per sampling location ranged from 12-29.

Note 1: Supplementary data and results

Supplementary raw data obtained from the analysis of samples collected from the large distribution (site 1) and reticulation network sampling points (site 2) refer to Tables A1-A3 in **Appendix A** of this report.

2.3.1.1 Phyla diversity

For the large drinking water distribution points, Proteobacteria is clearly seen as the dominant phyla. In sample Ben-Res2-IN (Figure 2.2), the phyla Proteobacteria, Bacteroidetes and Cyanobacteria were present across most of the time points. At five time points Verrucomicrobia was present at distinct abundance levels. In November of 2018, the Proteobacteria dominated the sample consisting of 99 %. The Brak-RS3-IN (Figure 2.3) samples had Proteobacteria and Bacteroidetes in common. Verrucomicrobia was also present at specific points but not at identical times compared to Ben-Res2-IN.

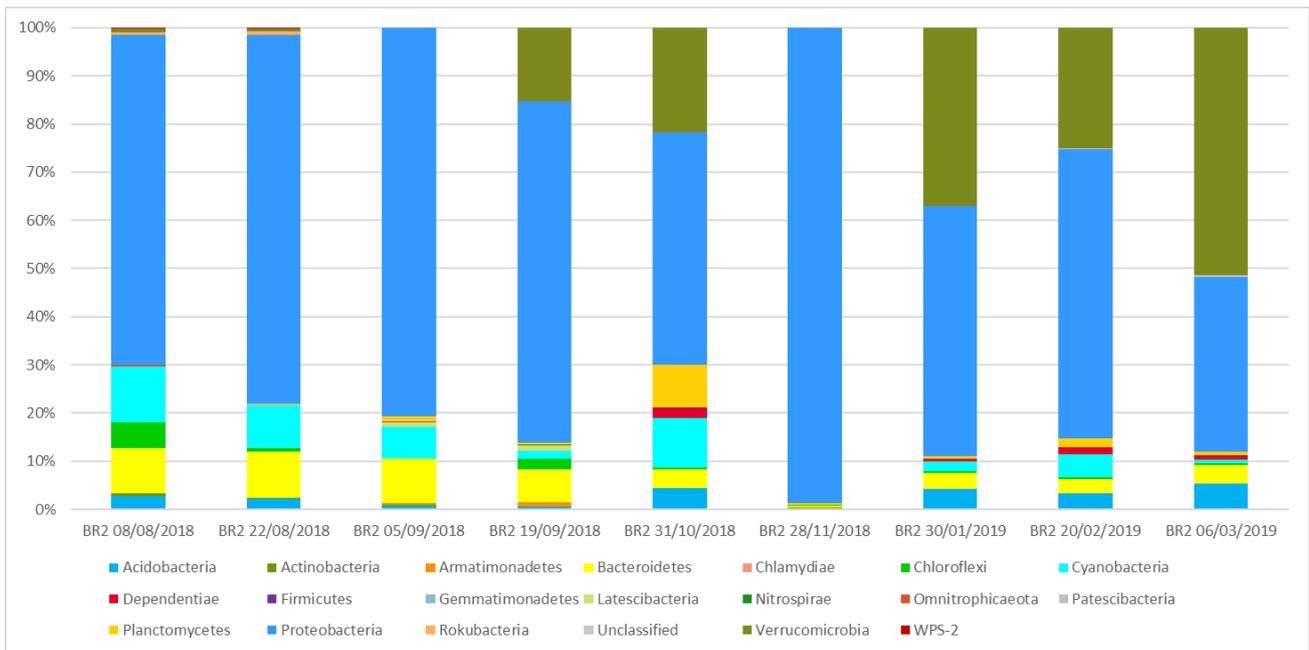


Figure 2.2: Relative abundance plot per phyla for sample location Ben-Res2-IN over the sampling period.

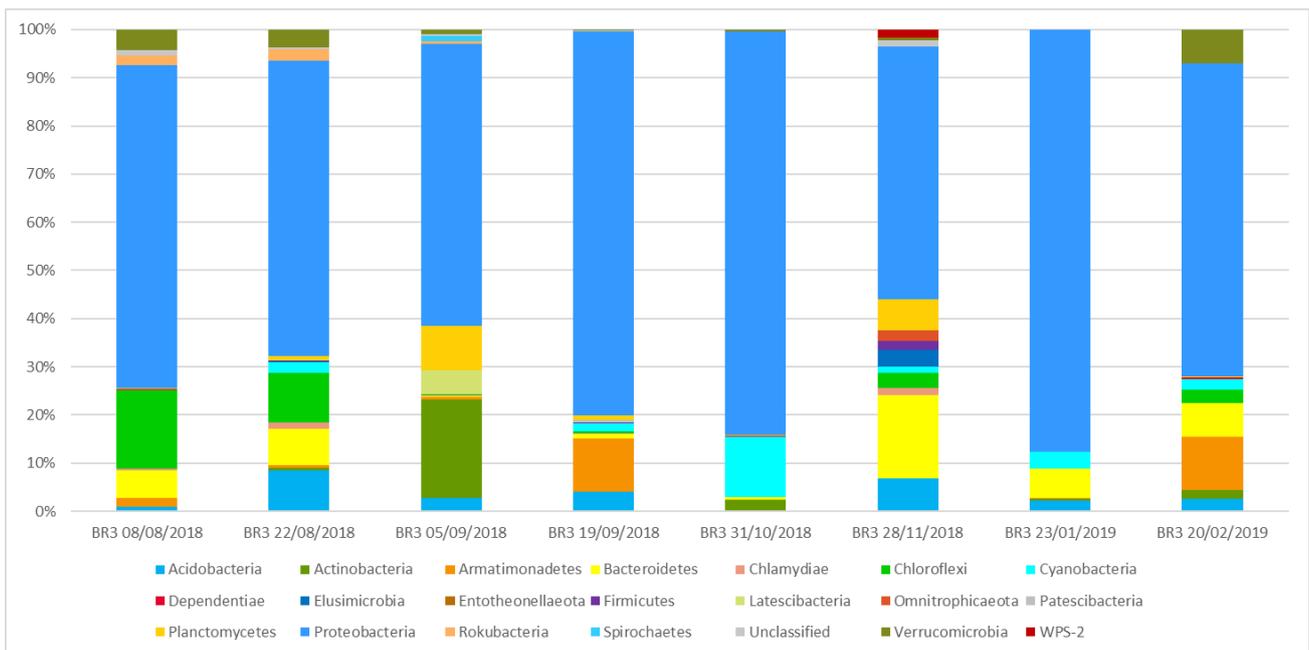


Figure 2.3: Relative abundance plot per phyla for sample location Brak-RS3-IN over the sampling period.

There were 12 unique phyla that were highly diverse at this location during November. The points Ben-Res2-IN and Brak-RS3-IN were obtained along the same route. Ga-Luka (Figure 2.4) also had the same common phyla across all sampling points as the first two points with the addition of Acidobacteria. Each sample however had unique phyla such as Latescibacteria; Armatimonadetes; Elusimicrobia; Planctomycetes; Firmicutes and Verrucomicrobia present at low percentages. The post-production point MB2 (Figure 2.5) had its own unique phyla apart from Proteobacteria being dominant in most of the samples.

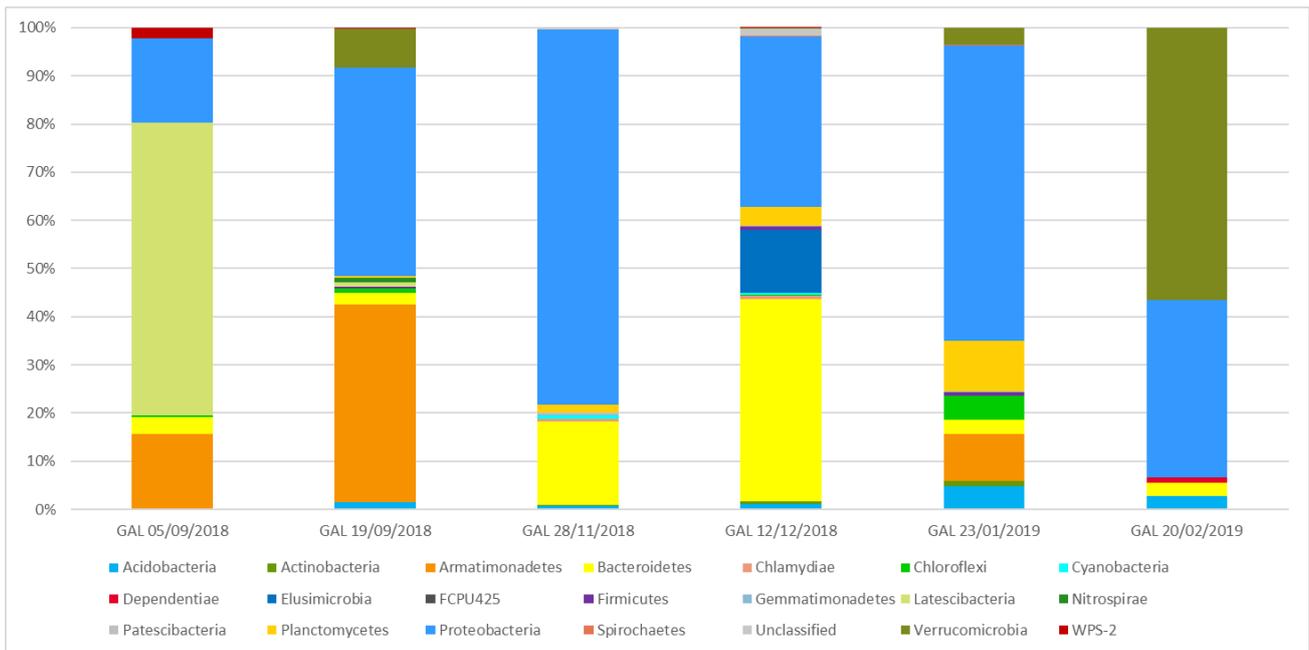


Figure 2.4: Relative abundance plot per phyla for sample location Ga-Luka over the sampling period.

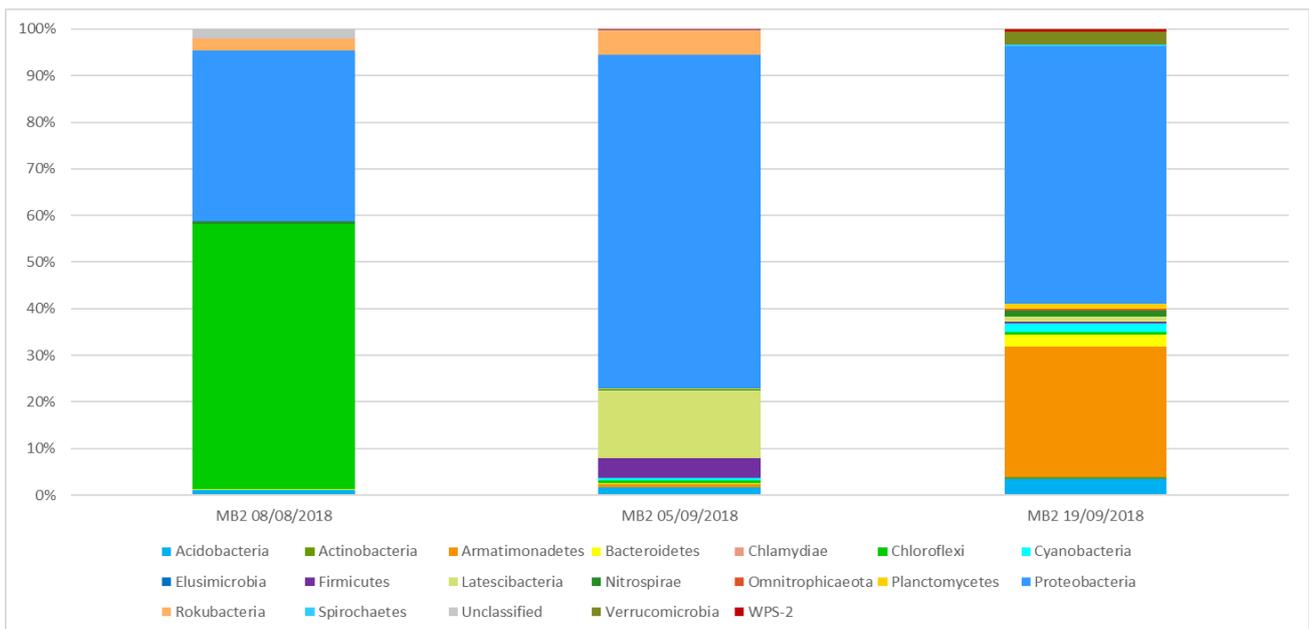


Figure 2.5: Relative abundance plot per phyla for sample location MB2 over the sampling period.

At the MB2 sampling location, Chloroflexi, Latescibacteria and Armatimonadetes was seen at a high percentage but in different samples, respectively. A high number of different phyla was present in samples taken at this location. P4PL (Figure 2.6), like the previous points, was dominated by Proteobacteria and additionally Acidobacteria, Bacteroidetes, Cyanobacteria and Planctomycetes were observed across all points. Unique phyla such as Firmicutes, Armatimonadetes and Dependientiae were also present at specific time points. This sampling location has the most shared phyla across all points. Rust-P6 (Figure 2.7) which was collected along the same route as Ga-Luka shared Acidobacteria and Actinobacteria in addition to Proteobacteria. Each of the time points had a unique phyla distribution as well with the presence of Zixibacteria and Omnitrophicaeota which were unique to this distribution sample location.

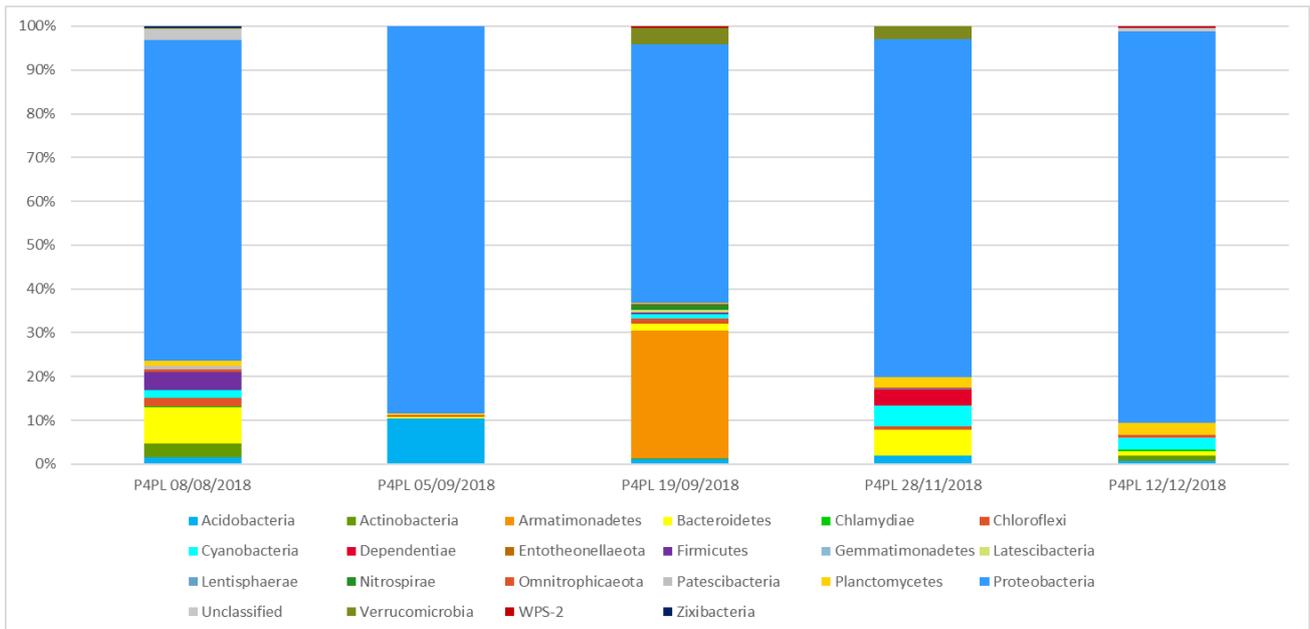


Figure 2.6: Relative abundance plot per phyla for sample location P4PL over the sampling period.

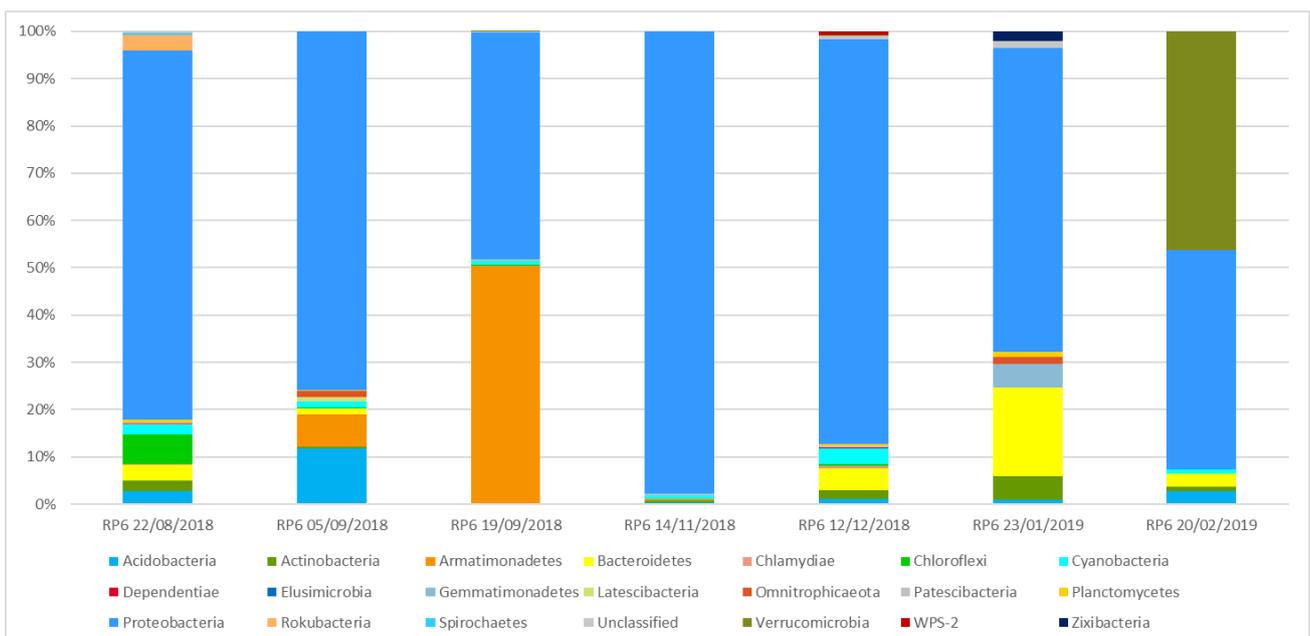


Figure 2.7: Relative abundance plot per phyla for sample location Rust-P6 over the sampling period.

The reticulation/ point-of-use samples were collected from six different points and each location had a varying number of samples which could be used for this study. Proteobacteria was the most abundant phylum in most of the samples sequenced. AgricBuild (Figure 2.8) had a large phylum diversity amongst the samples with the majority of the phyla shared amongst samples. The December sample was dominated by the Proteobacteria (91%) and did not share phyla such as Cyanobacteria; Omnitrophicaeota; Spirochaetes and Verrucomicrobia. The different samples had varying percentages of each phyla and phyla such as Omnitrophicaeota and Elusimicrobia were present at larger percentages in two samples.

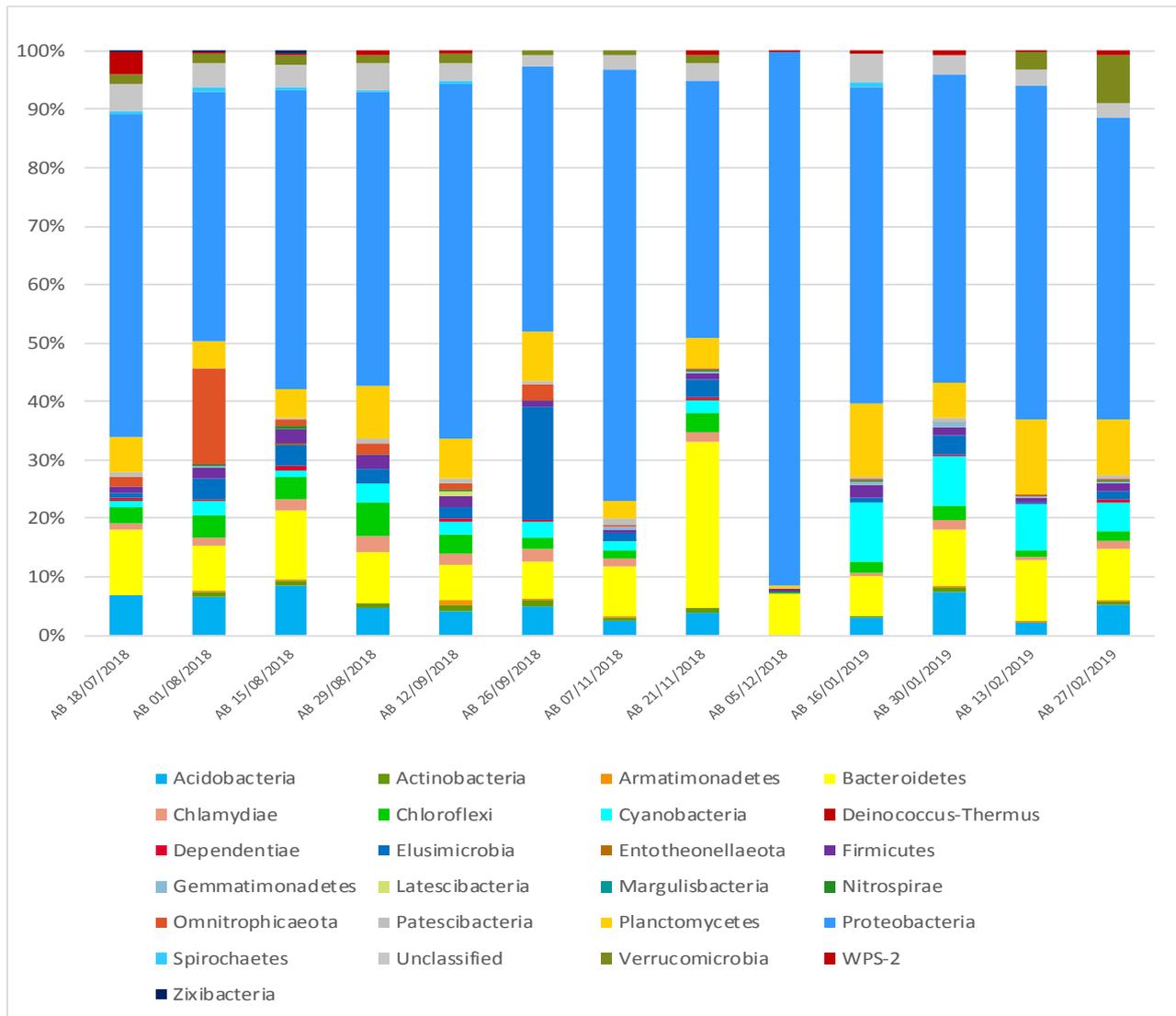


Figure 2.8: Relative abundance plot per phyla for sample location AgricBuild over the sampling period.

Groenkloof (Figure 2.9) had a unique diversity of phyla in each sample. The November samples were dominated by the Proteobacteria at 97% whereas the other time points had higher diversity with varying abundance levels observed for each sample.

Location NS2 (Figure 2.10) had a more stable community dominated by the Proteobacteria. Verrucomicrobia, Planctomycetes, Chloroflexi, Actinobacteria and Chlamydia were present at higher abundances in specific samples. Most of the phyla were shared amongst all samples apart from Acidobacteria and Firmicutes which were present in only two samples.

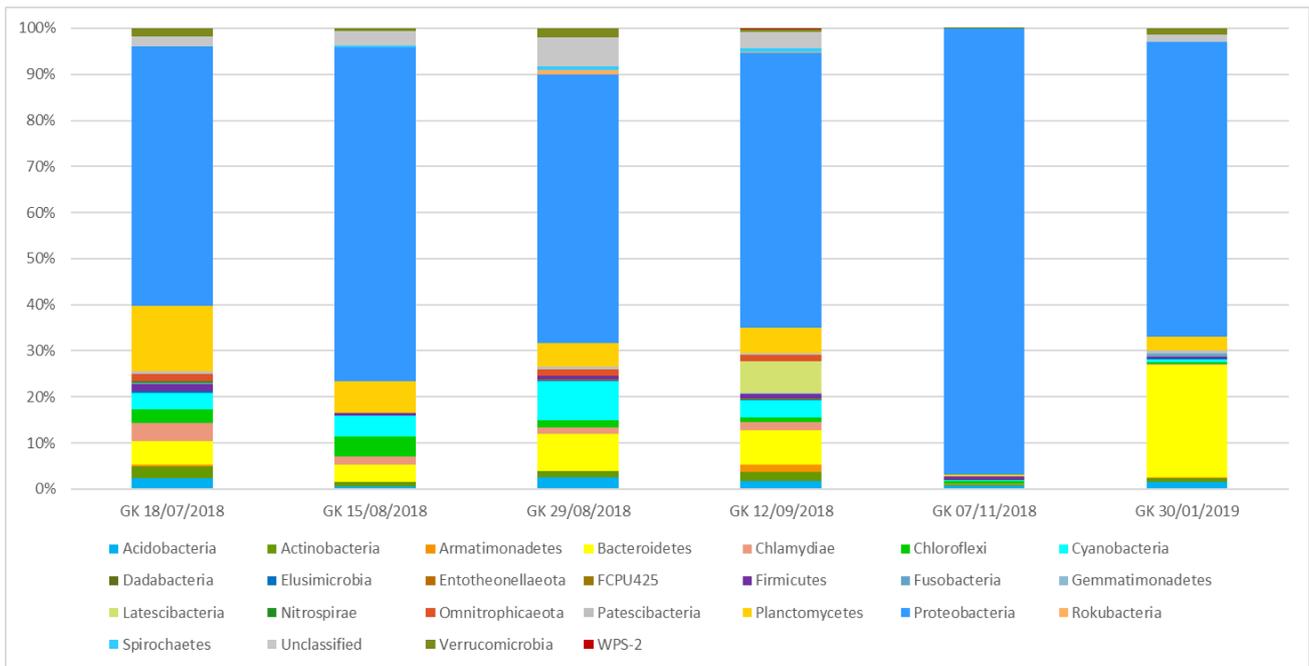


Figure 2.9: Relative abundance plot per phyla for sample location Groenkloof over the sampling period.

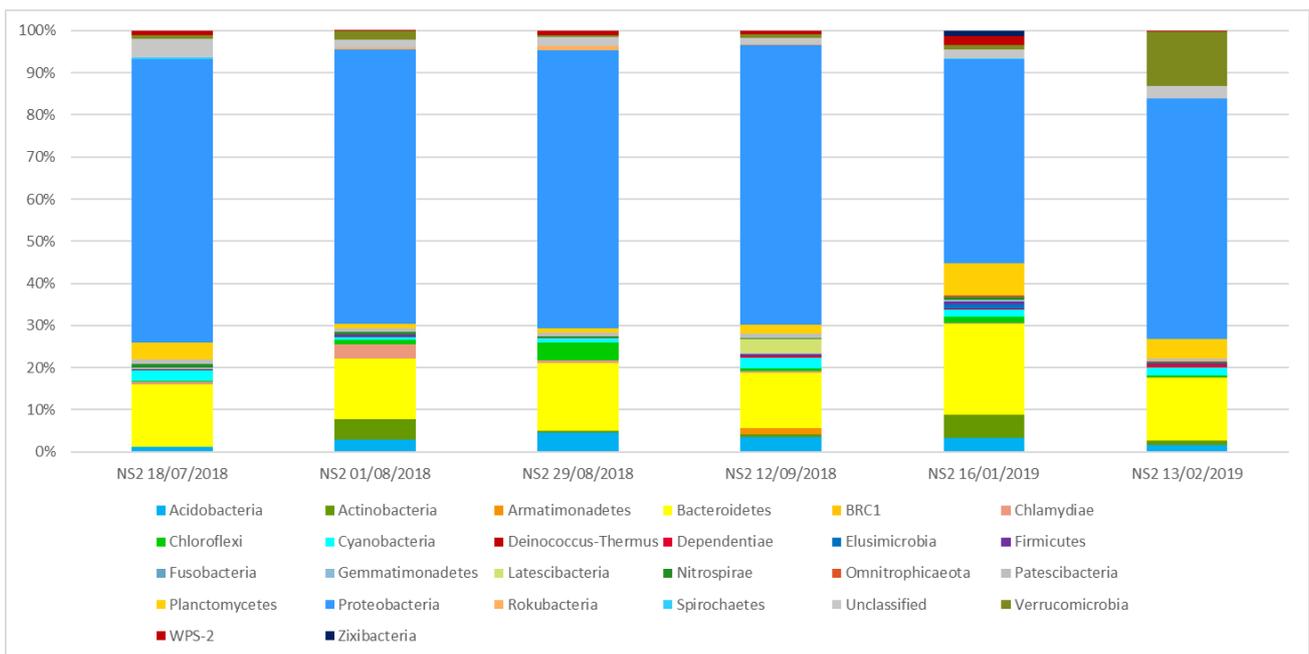


Figure 2.10: Relative abundance plot per phyla for sample location NS2 over the sampling period.

Silverton (Figure 2.11) was highly diverse and each sample contained a number of different phyla. The phyla Rokubacteria and Patescibacteria were present in two of the samples taken at this sample location. Although the Proteobacteria also dominated in some of the Valhalla samples (Figure 2.12) a number of other phyla were also present at higher abundances. This was different from what was observed for most of the other sampling points. The phyla present included Chloroflexi, Bacteroidetes, Acidobacteria and Cyanobacteria. The phylum Gemmatimonadetes was unique to Valhalla and Rokubacteria (which was only shared with Silverton) was also observed.

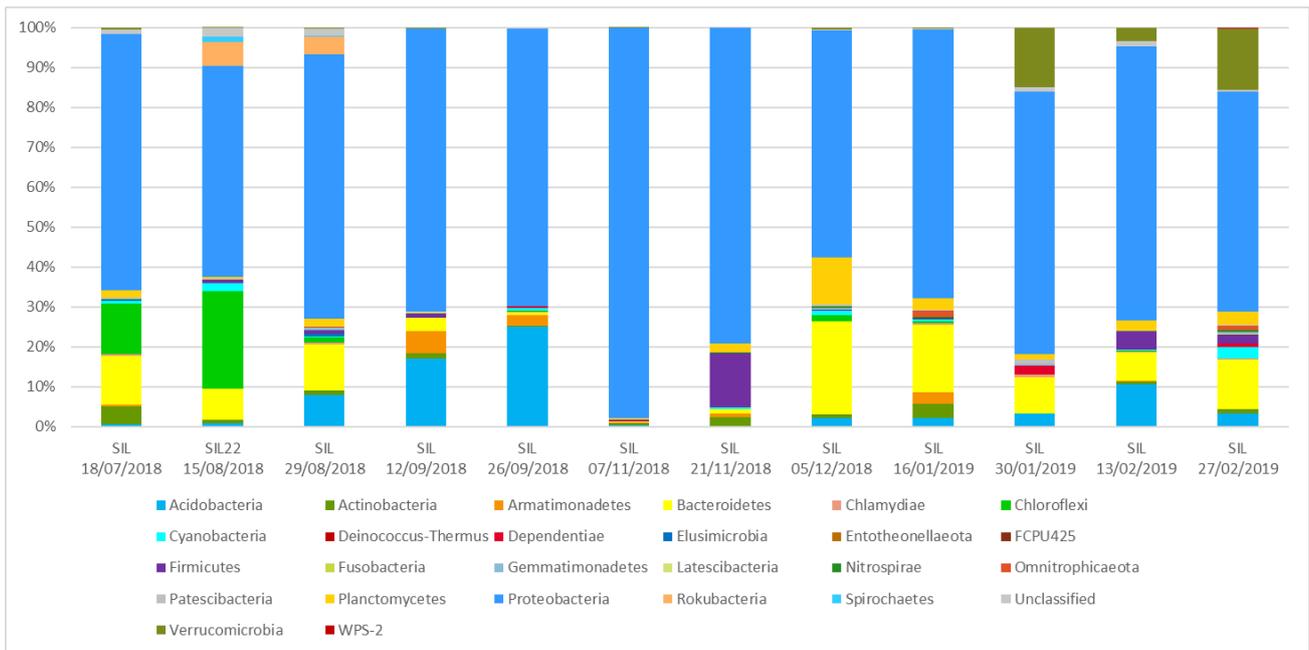


Figure 2.11: Relative abundance plot per phyla for sample location Silvertown over the sampling period.

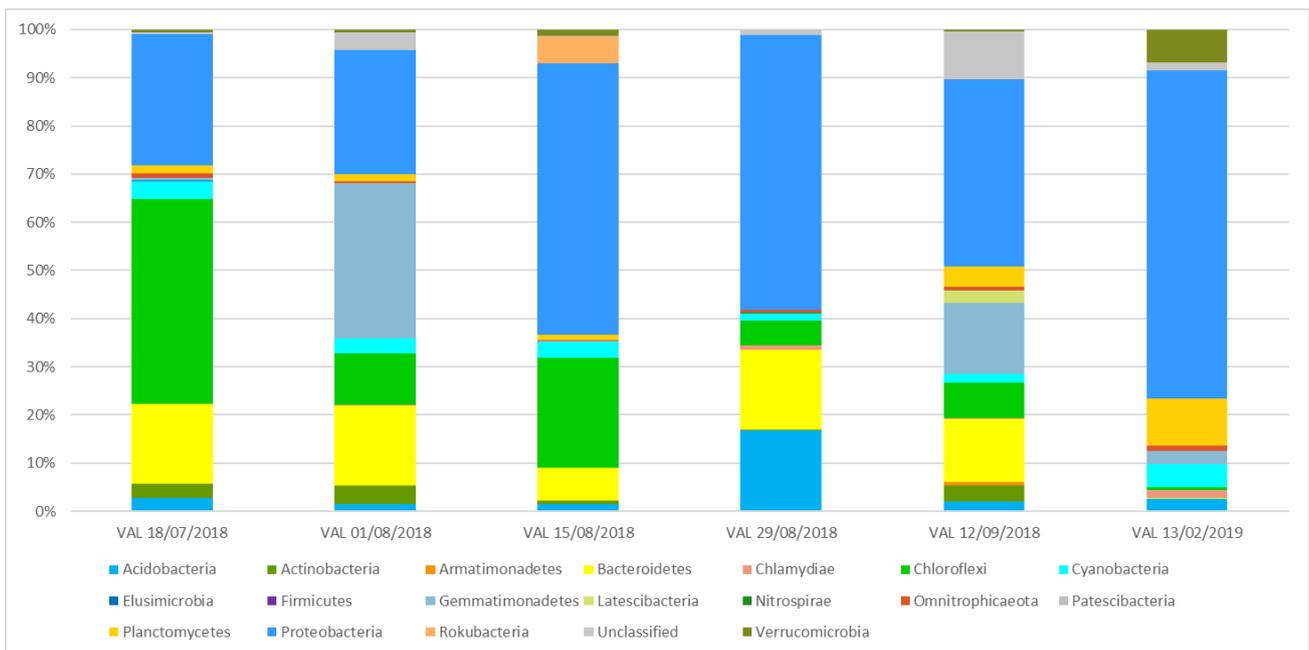


Figure 2.12: Relative abundance plot per phyla for sample location Valhalla over the sampling period.

Waverley (Figure 2.13) had diverse and unique communities present in all the samples. The phyla present were detected in almost all the samples from this location but some samples had a unique distribution of phyla. The phyla Armatimonadetes, Planctomycetes, Spirochaetes, Proteobacteria and Chloroflexi were present at higher relative abundances at this location. An interesting trend observed was that phyla seen as dominant at one specific sampling time point were often rare or absent from another sample taken from the same sampling location. The sampling locations had very different bacterial communities at the different times when samples were collected. This was observed at all the sampling locations.

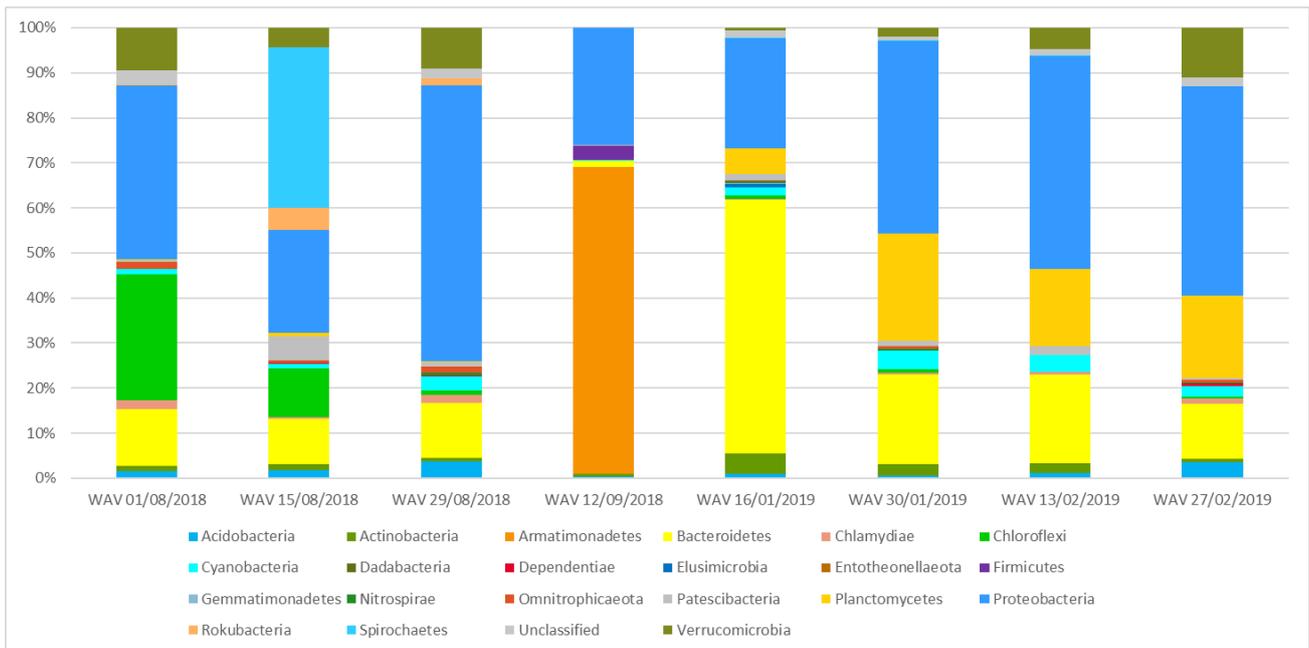


Figure 2.13: Relative abundance plot per phyla for sample location Waverley over the sampling period.

2.3.1.2 Class diversity

In the large distribution network (Site 1) dataset, the dominant classes varied across each time point and the overall most abundant class was analysed in greater detail. For most of the samples from Ben-Res2-IN (Figure 2.14) the Alpha- and Gammaproteobacteria had a similar abundance. In one sample however the Gammaproteobacteria had a relative abundance of 99%. The Deltaproteobacteria were also detected in one sample.

Brak-RS3-IN (Figure 2.15) was also dominated by Alpha- and Gammaproteobacteria with one sample having Deltaproteobacteria present.

Sample location Ga-Luka (Figure 2.16) was also dominated by Alpha- and Gammaproteobacteria. In one sample the Alphaproteobacteria dominated with 97% and in another sample Deltaproteobacteria were detected.

The production point MB2 (Figure 2.17) was almost completely dominated by Gammaproteobacteria with a small presence of Alphaproteobacteria.

Location P4PL (Figure 2.18) had three samples dominated by Alphaproteobacteria and two samples dominated by the Gammaproteobacteria at a high relative abundance.

Rust-P6 (Figure 2.19) was also dominated by Alpha- and Gammaproteobacteria. In the colder months, the samples were dominated by Gammaproteobacteria thereafter it changed with the Alphaproteobacteria dominating.

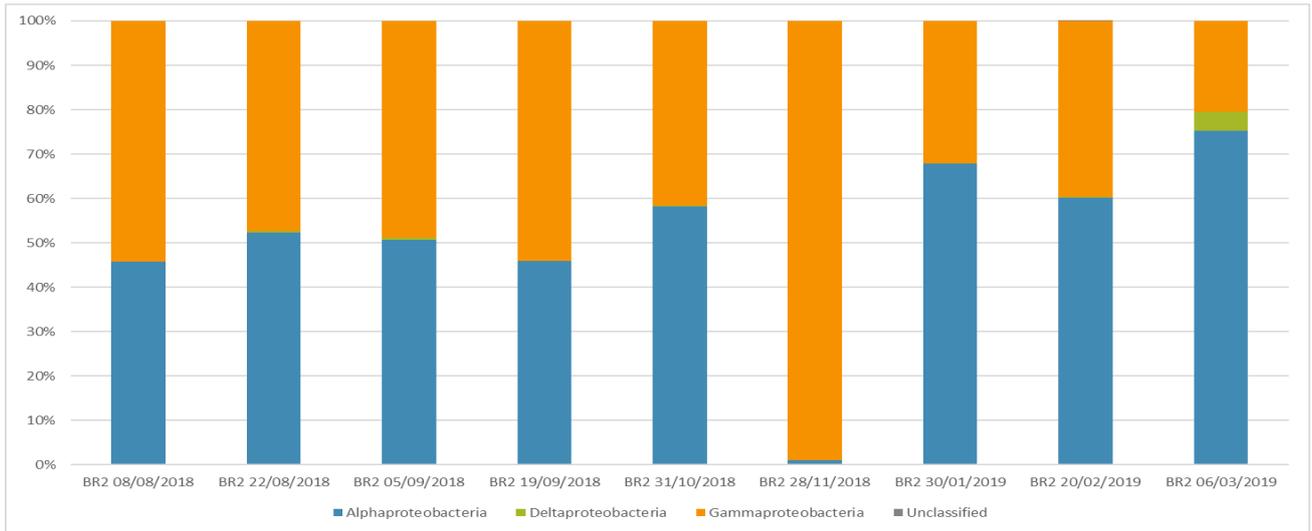


Figure 2.14: Relative abundance plot for the classes in phylum Proteobacteria for sample location Ben-Res2-IN over the sampling period.

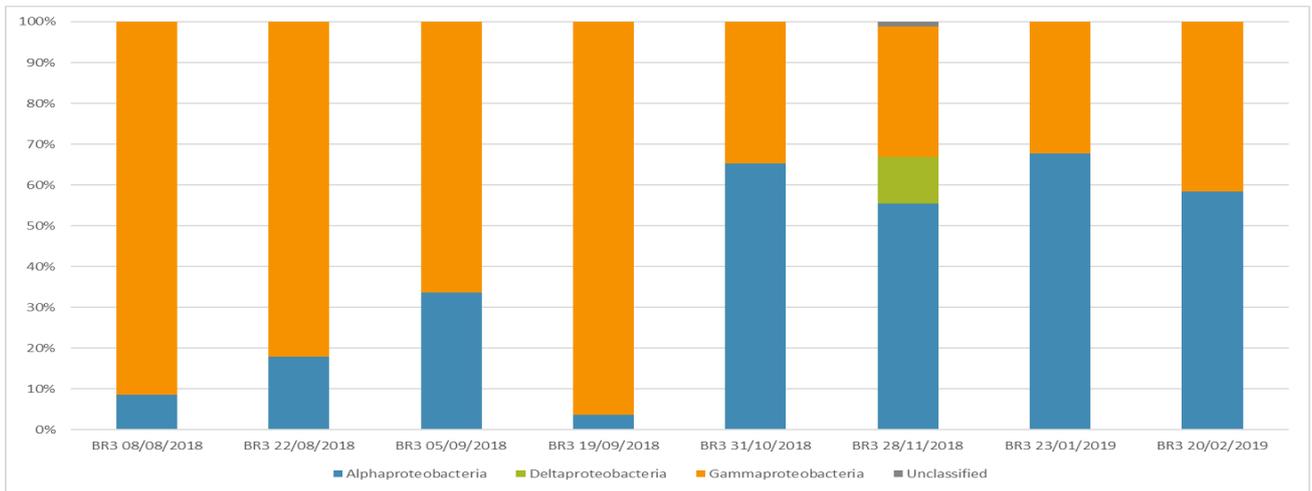


Figure 2.15: Relative abundance plot for the classes in phylum Proteobacteria for sample location Brak-RS3-IN over the sampling period.

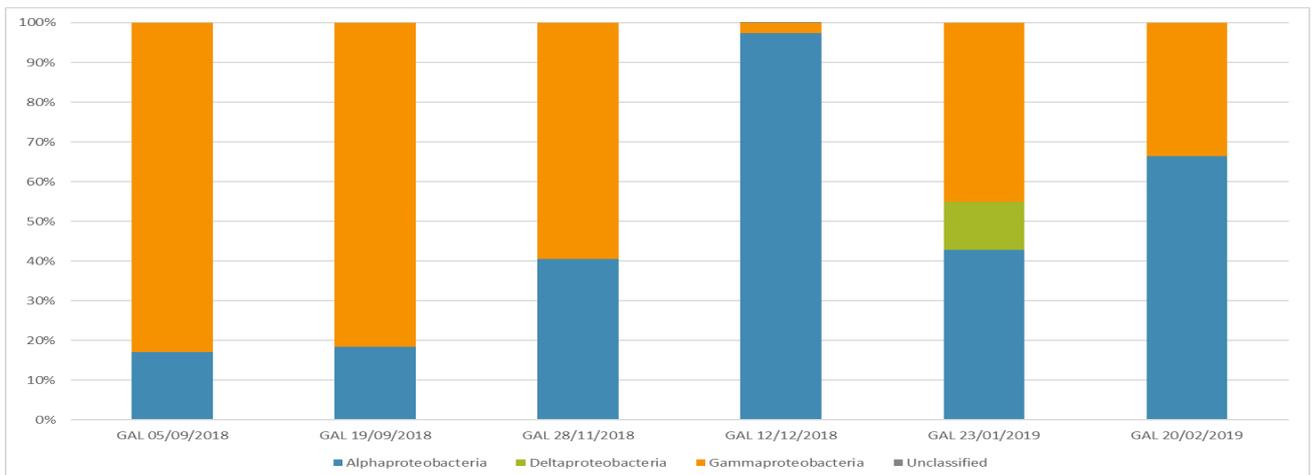


Figure 2.16: Relative abundance plot for the classes in phylum Proteobacteria for sample location Ga-Luka over the sampling period.

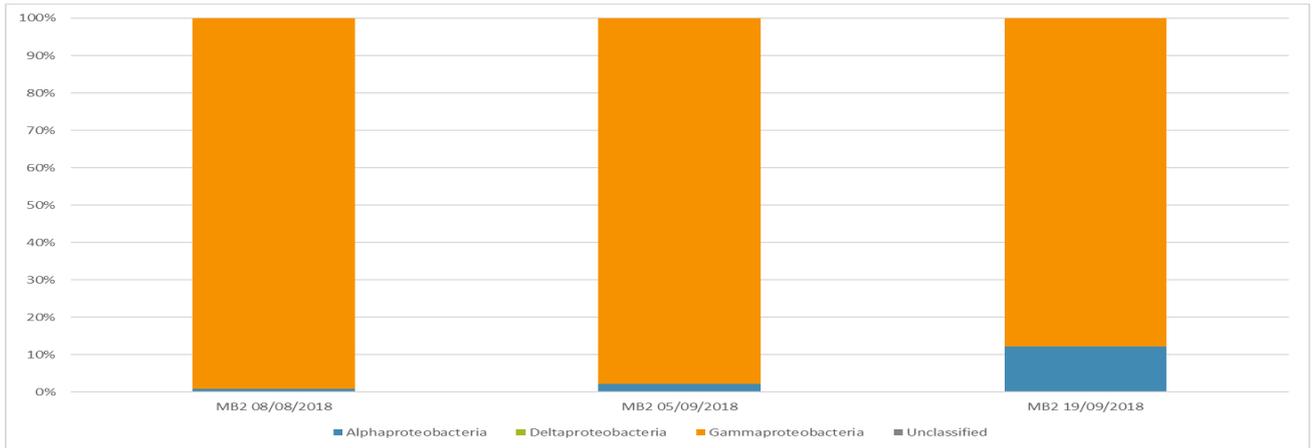


Figure 2.17: Relative abundance plot for the classes in phylum Proteobacteria for sample location MB2 over the sampling period.

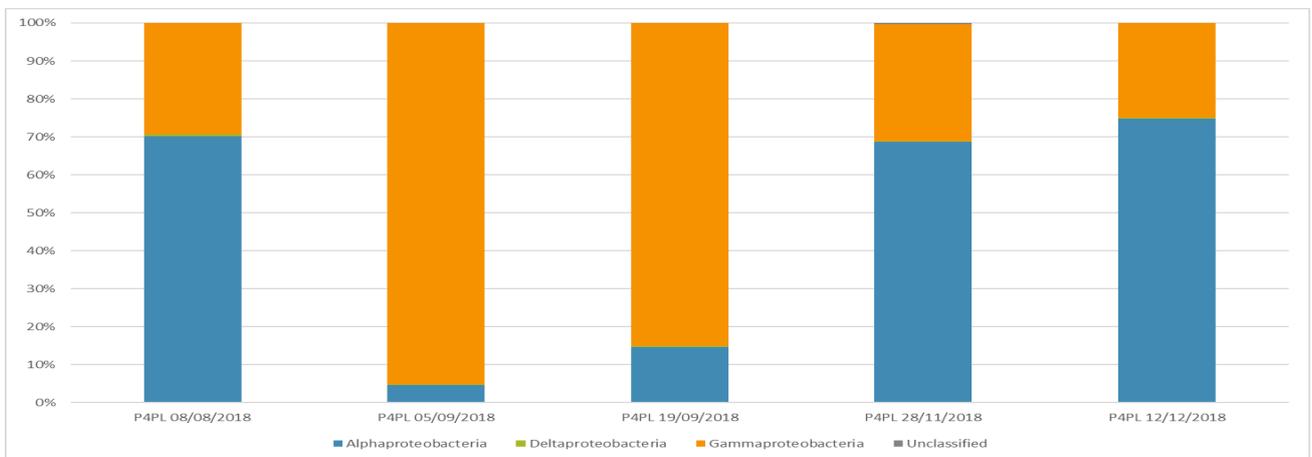


Figure 2.18: Relative abundance plot for the classes in phylum Proteobacteria for sample location P4PL over the sampling period.

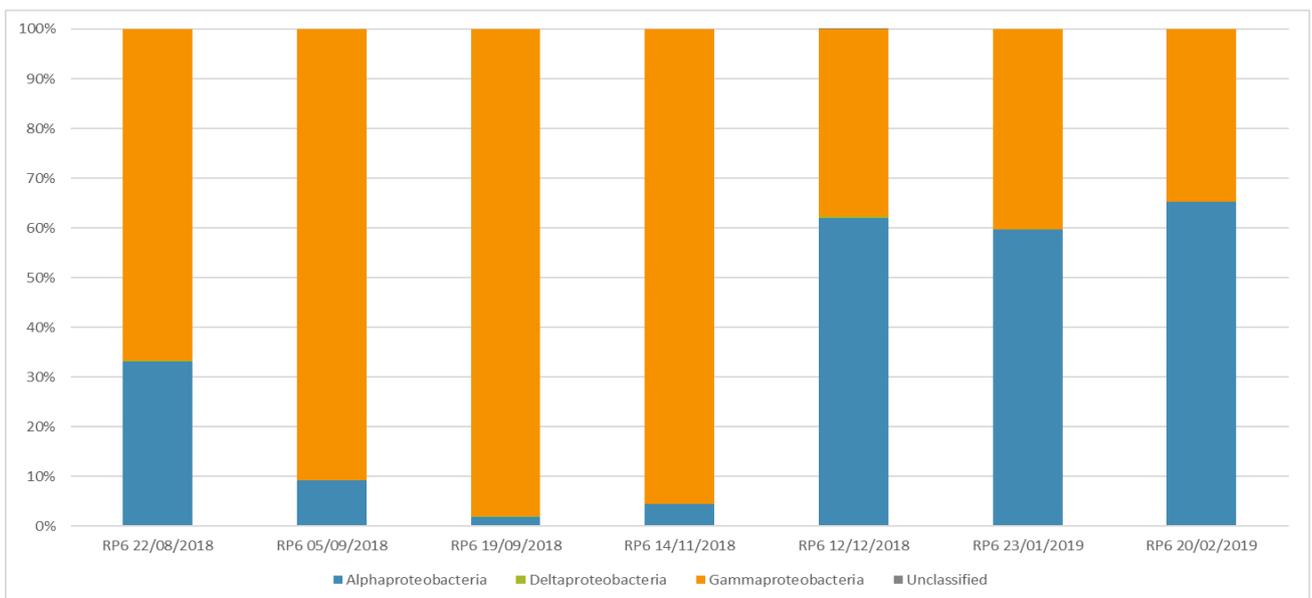


Figure 2.19: Relative abundance plot for the classes in phylum Proteobacteria for sample location RP6 over the sampling period.

In the reticulation sample locations, the distribution of the classes differed between locations. The samples were either dominated by the Alpha- or Gammaproteobacteria and at four of the six locations the Deltaproteobacteria were also regularly detected but they never dominated the system. AgricBuild (Figure 2.20) had all three classes present in twelve of the thirteen samples analysed with an overall dominance in Gammaproteobacteria. Groenkloof (Figure 2.21) had two samples which were dominated by only Alpha- and Gammaproteobacteria. The other samples at this location had a small percentage of Deltaproteobacteria present.

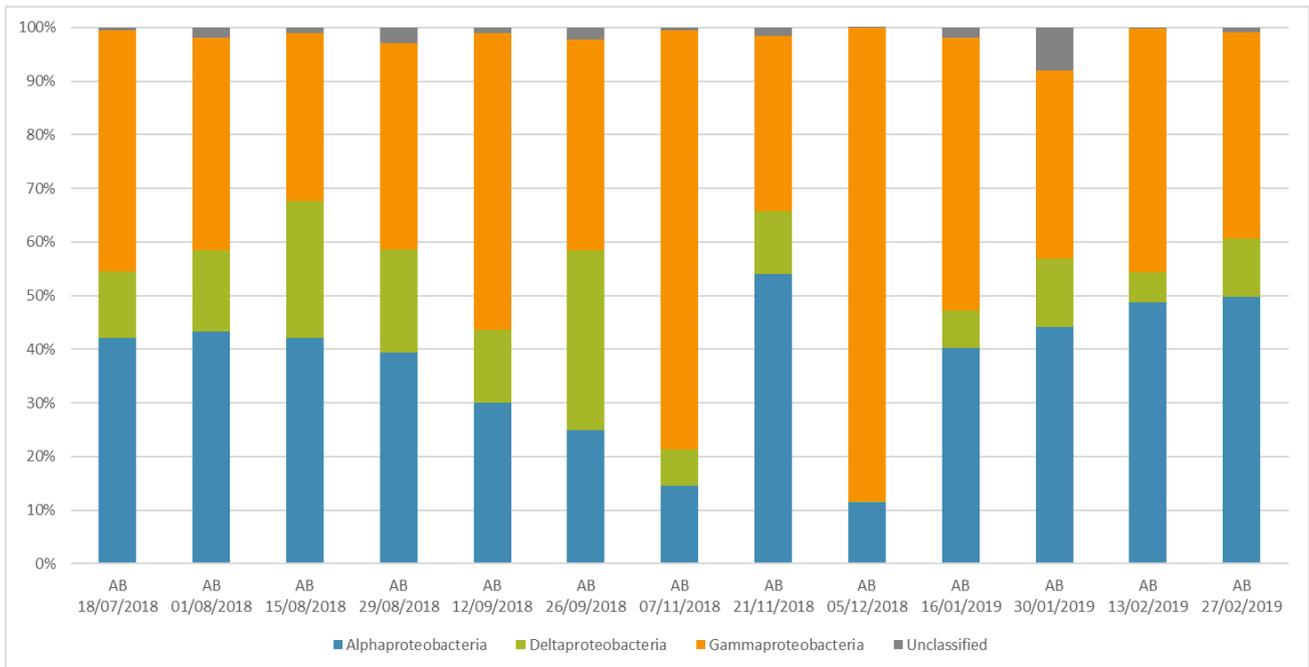


Figure 2.20: Relative abundance plot for the classes in phylum Proteobacteria for sample location AgricBuild over the sampling period.

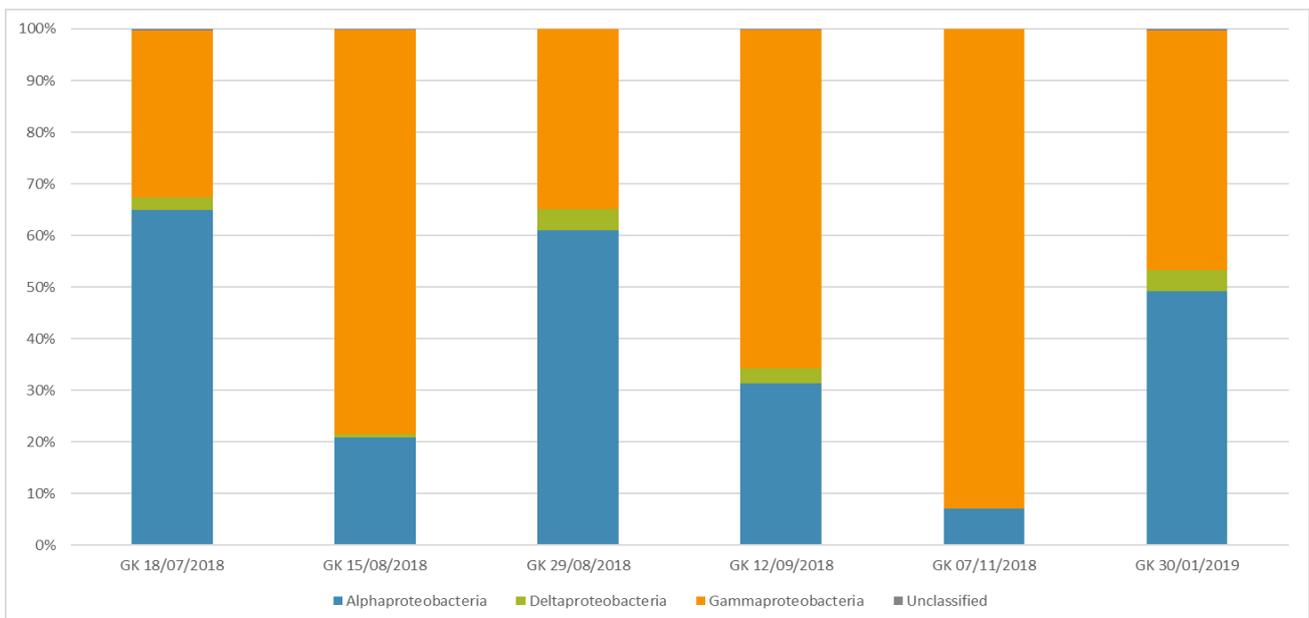


Figure 2.21: Relative abundance plot for the classes in phylum Proteobacteria for sample location Groenkloof over the sampling period.

Sample location NS2 (Figure 2.22) was overall dominated by the Alphaproteobacteria and Gamma- and Deltaproteobacteria were present at lower abundances. In the Silverton (Figure 2.23) samples, an increase in the Gammaproteobacteria was seen at a few time points and then an increase in the Gammaproteobacteria with the Deltaproteobacteria appearing towards the end of the sampling period. Valhalla (Figure 2.24) in contrast to Silverton were largely dominated by the Alphaproteobacteria with only two samples where the Gammaproteobacteria dominated. The location Waverley (Figure 2.25) also had samples dominated by either the Alpha- or Gammaproteobacteria. The Deltaproteobacteria were present at a low relative abundance in some of the samples.

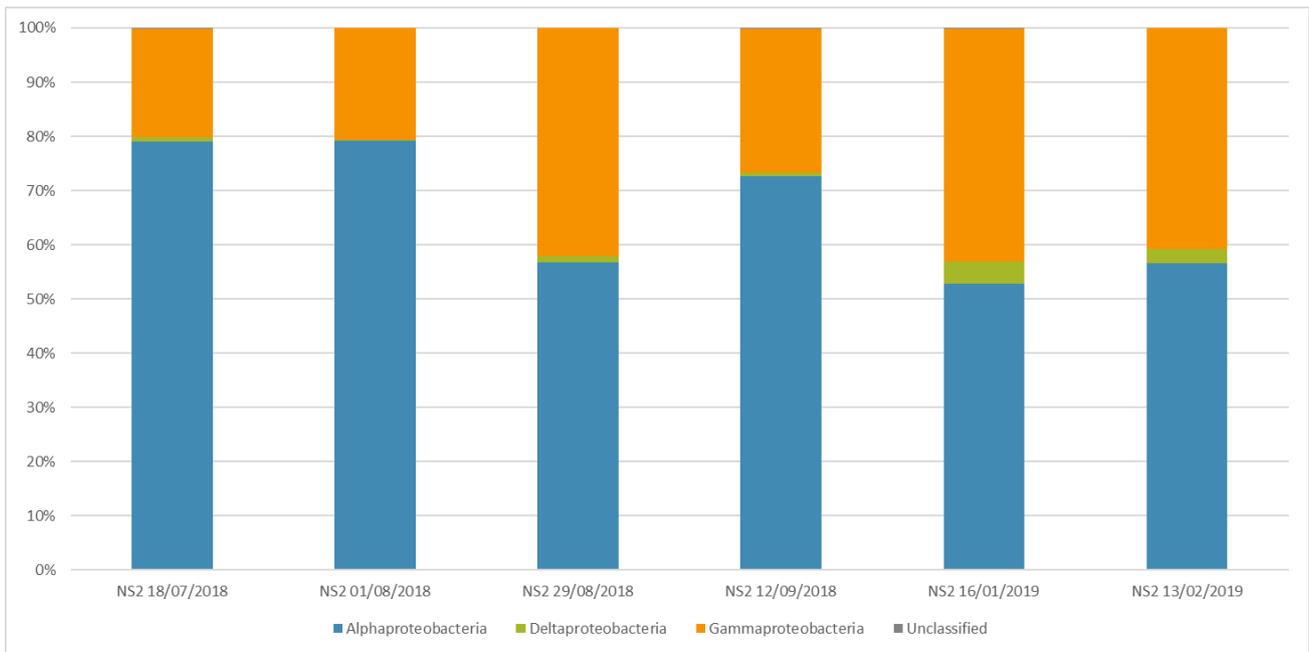


Figure 2.22: Relative abundance plot for the classes in phylum Proteobacteria for sample location NS2 over the sampling period.

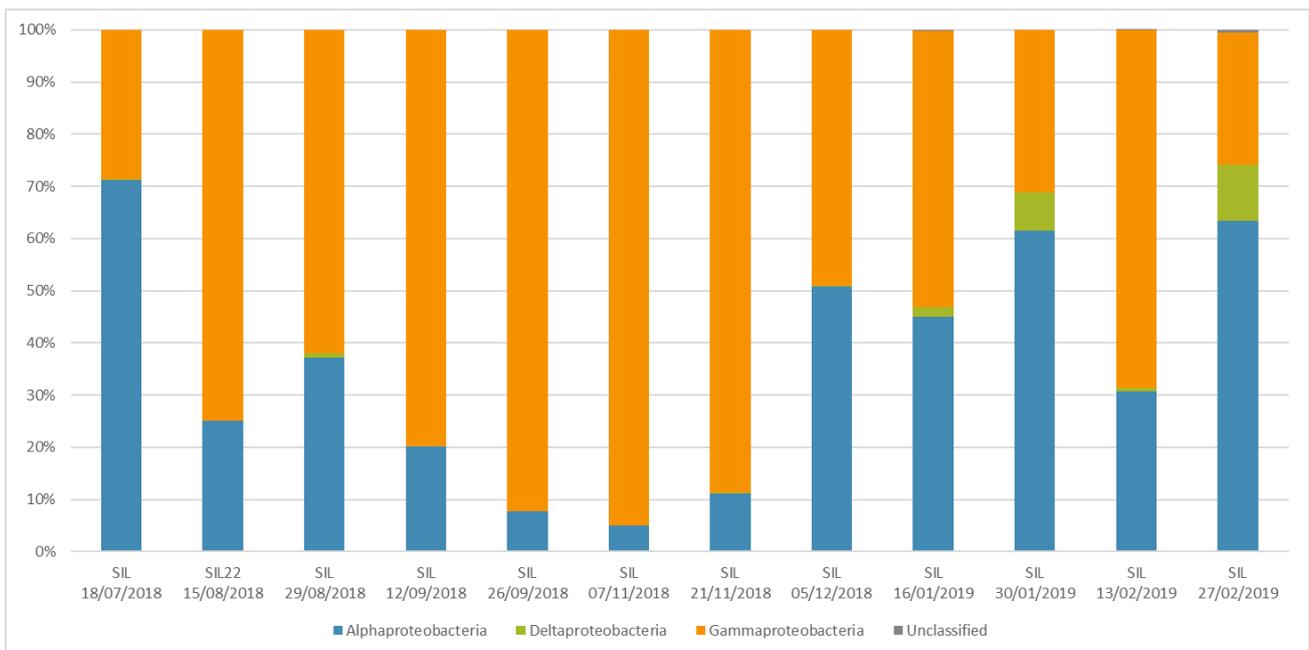


Figure 2.23: Relative abundance plot for the classes in phylum Proteobacteria for sample location Silverton over the sampling period.

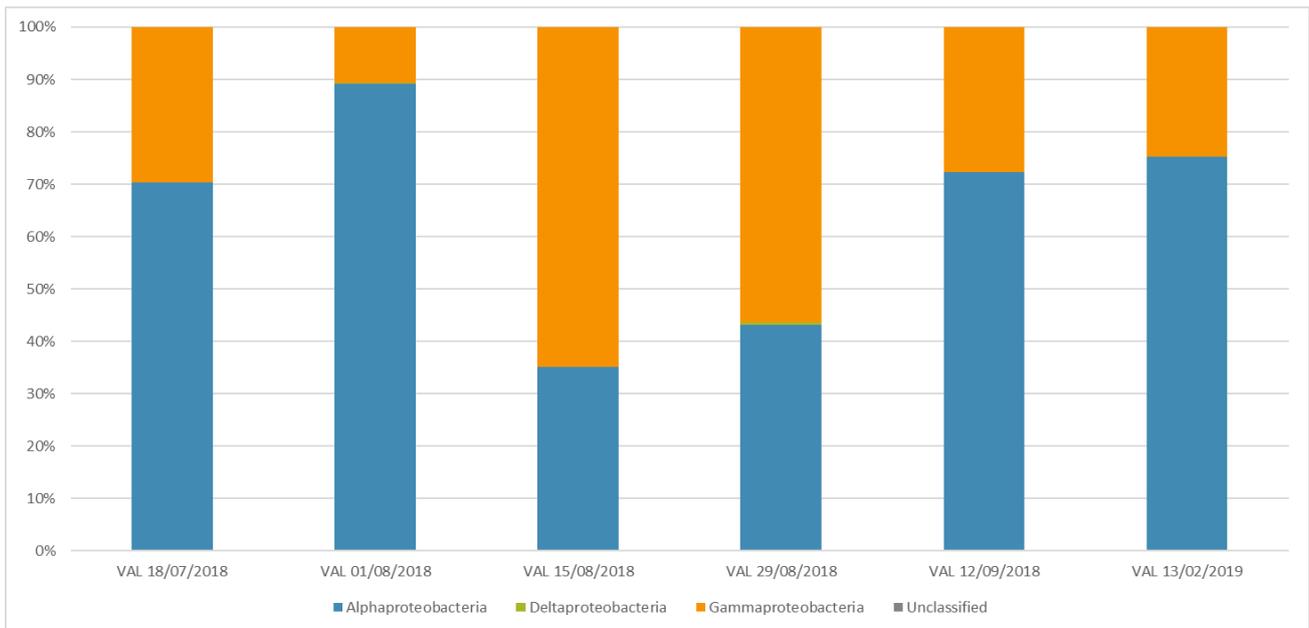


Figure 2.24: Relative abundance plot for the classes in phylum Proteobacteria for sample location Valhalla over the sampling period.

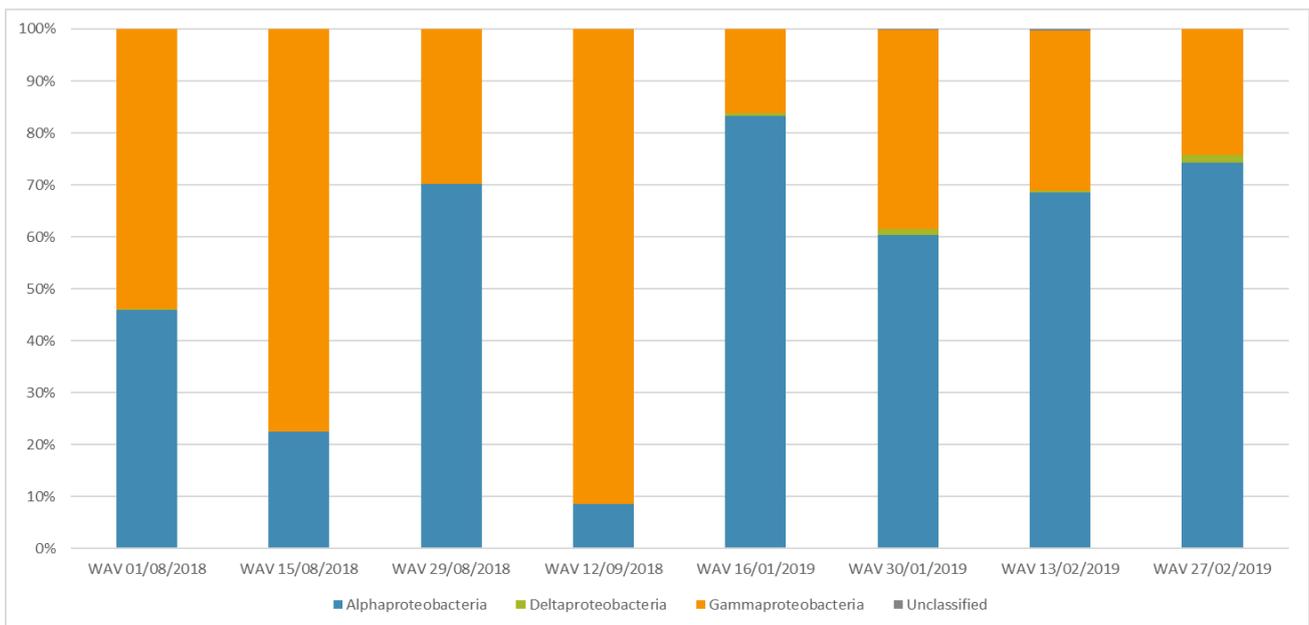


Figure 2.25: Relative abundance plot for the classes in phylum Proteobacteria for sample location Waverley over the sampling period.

2.3.1.3 Family diversity

In the distribution network, the community diversity varied between locations and samples from the same location. For Ben-Res2-IN (Figure 2.26), the class Alphaproteobacteria dominated with the *Sphingomonadaceae*, the only family shared amongst all points at this location. The families which were shared between specific samples were *Hyphomonadaceae*; *Devosiaceae*; *Xanthobacteraceae* and *Caulobacteraceae*. The November sample differed significantly from other samples with the *Caulobacteraceae* being the dominant family. This could have been influenced by changes at the sample location that may have occurred as a result of treatment, contamination or differences in abiotic factors. The

overall composition of samples taken before and after this sample were similar in diversity. The Brak-RS3-IN (Figure 2.27) samples were dominated by the class Gammaproteobacteria and the *Burkholderiaceae*, shared amongst all the points, was often the most dominant family. *Pseudomonadaceae* and *Nitrosomonadaceae* was also dominant in some of the samples taken at this location. Other families were present in the different samples but at a low relative abundance. Like Ben-Res2-IN, the November sample differed in diversity from all other points and had unique families like *Methylophilaceae* and *Acidiferrobacteraceae*. There were also two samples where the *Nitrosomonadaceae* was the dominant representative of the Gammaproteobacteria.

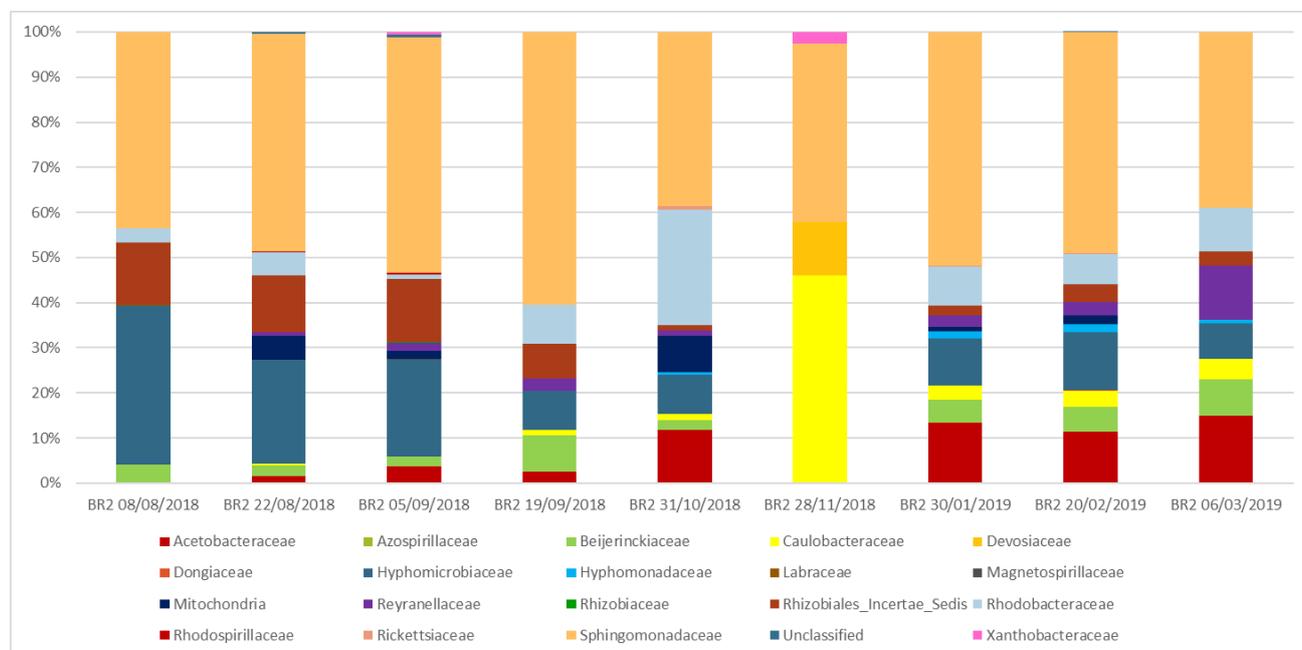


Figure 2.26: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Ben-Res2-IN over the sampling period.

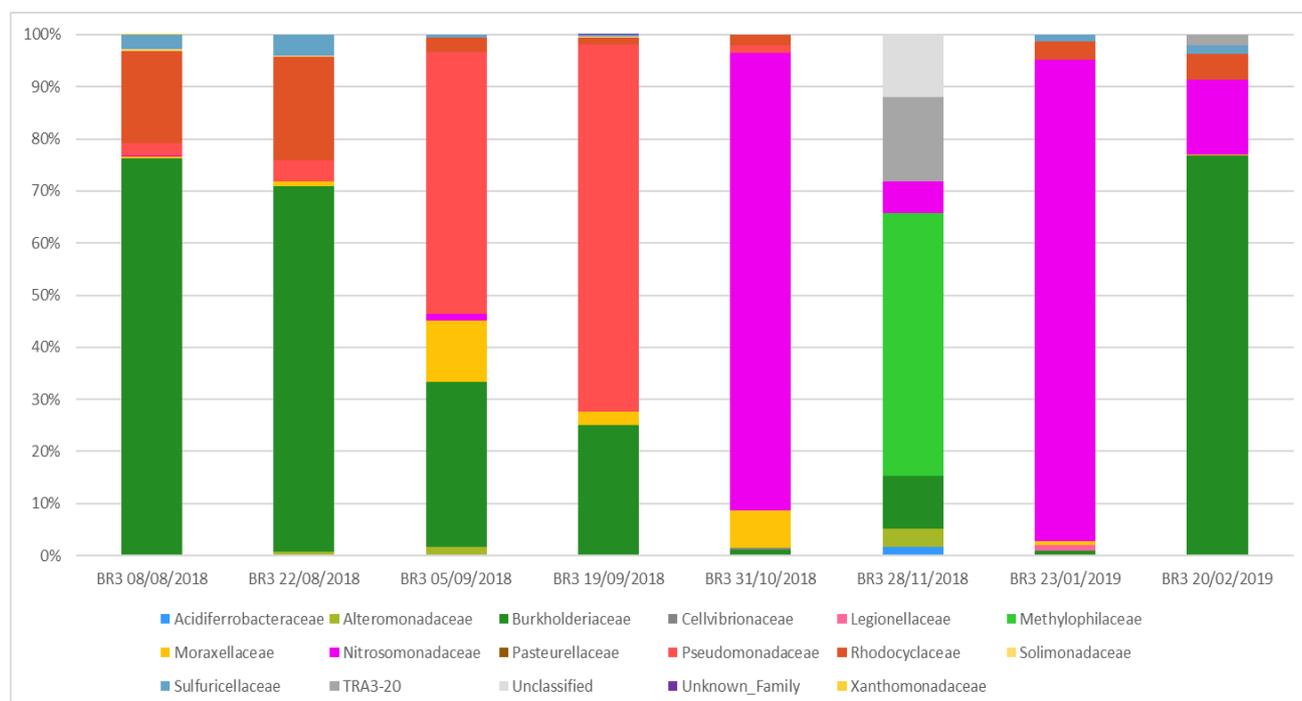


Figure 2.27: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Brak-RS3-IN over the sampling period.

The Ga-Luka sample (Figure 2.28), dominant in Gammaproteobacteria had significant diversity per sample location with *Burkholderiaceae* shared between all points. *Moraxellaceae* and *Pseudomonadaceae* were considerable dominant at this location and *Gallionellaceae* had a fairly high relative abundance in the December sample. The November sample, as was observed for the previous two locations, was also shown to be different from the other time points and *Moraxellaceae* had a 96% relative abundance among the Gammaproteobacteria. The production sample (Figure 2.29), which had only three samples that could be analysed for community diversity due to low DNA yield from the other samples had a unique abundance of families. *Burkholderiaceae* and *Pseudomonadaceae* were the families which were common between the points. *Rhodocyclaceae* was unique to the most diverse sample taken in September.

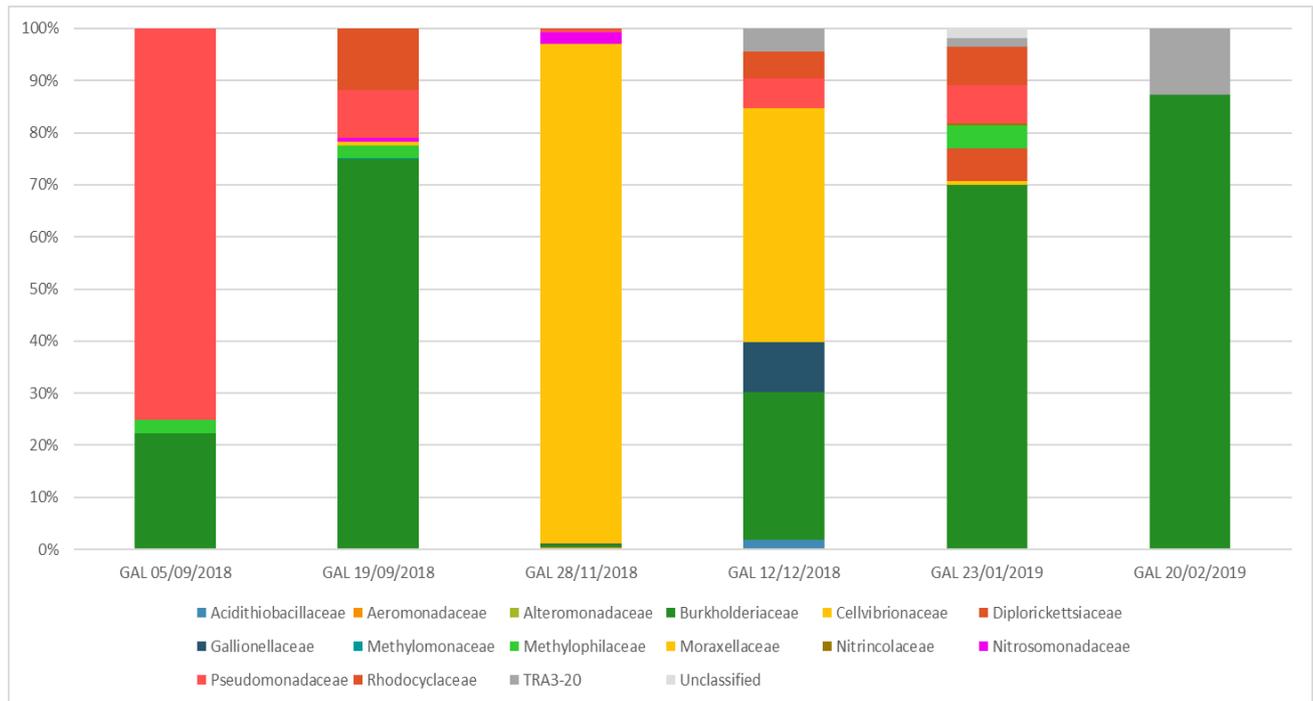


Figure 2.28: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Ga-Luka over the sampling period.

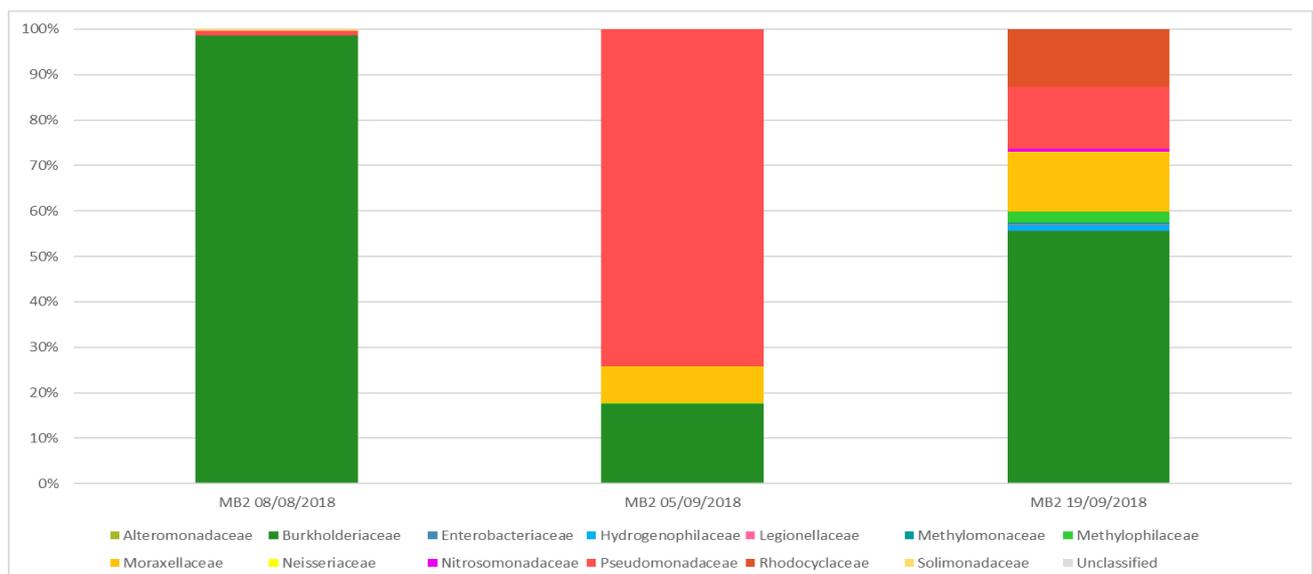


Figure 2.29: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample MB2 over the sampling period.

P4PL (Figure 2.30) was also diverse for all samples but had more families which were shared between all the samples. These families included: *Burkholderiaceae*, *Moraxellaceae*, *Nitrosomonadaceae* and *Pseudomonadaceae*. The point Rust-P6 (Figure 2.31) was also diverse and shared the same families amongst some samples as was observed for P4PL. *Alteromonadaceae* was unique in relative abundance to this sample and was observed only in one sample. The sample location had a high abundance of the same four families shared at points in P4PL, e.g. *Burkholderiaceae*, *Moraxellaceae*, *Nitrosomonadaceae* and *Pseudomonadaceae*. The November sample was clearly the distinct point amongst the samples.

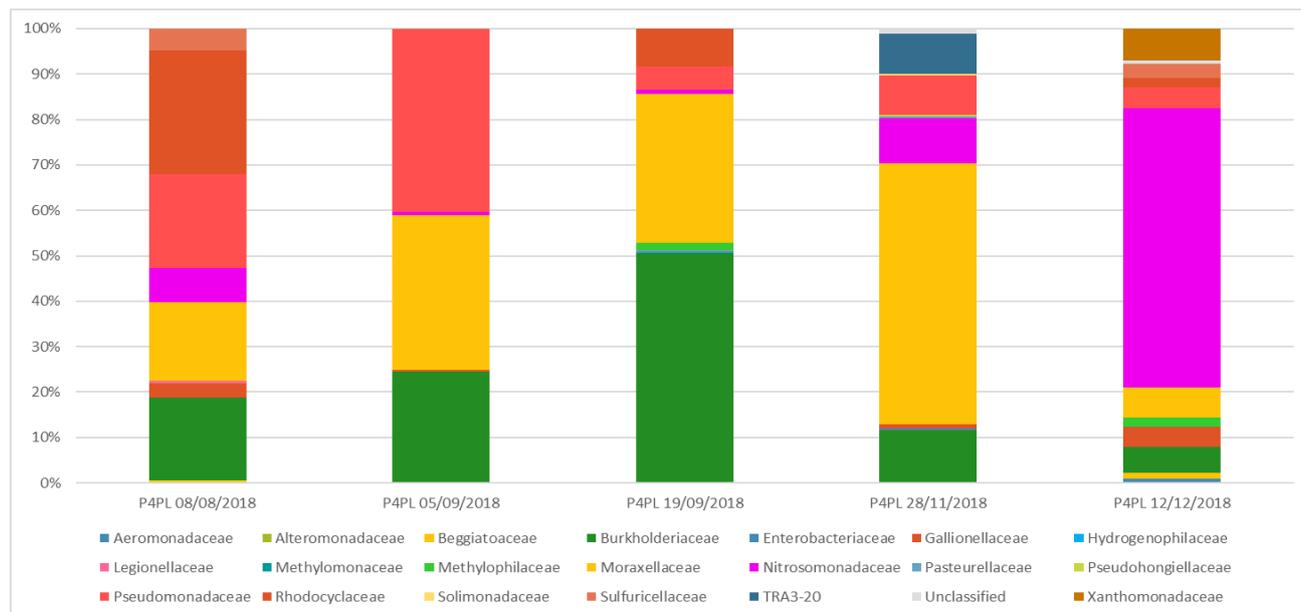


Figure 2.30: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample P4PL over the sampling period.

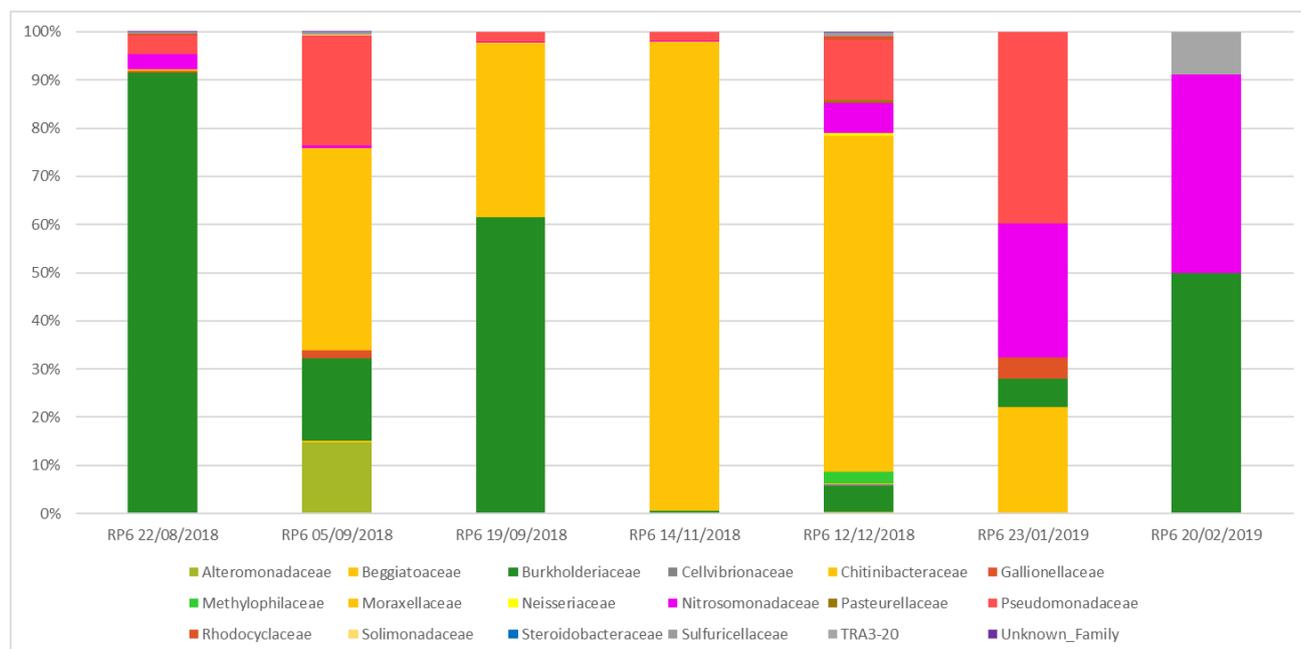


Figure 2.31: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample Rust-P6 over the sampling period.

For the reticulation system, more samples could be analysed due to the abundance of bacterial DNA obtained from the samples. In the AgricBuild (Figure 2.32) samples *Burkholderiaceae*, *Methylophilaceae* and *Nitrosomonadaceae* were shared amongst all sample locations. Three samples had a higher relative abundance of *Moraxellaceae* with two where relative abundances of higher than 75% were observed. Groenkloof (Figure 2.33) had a unique diversity across all samples and one outlier sample which was dominated by *Moraxellaceae* at 99%. Another sample had *Burkholderiaceae* dominating the Gammaproteobacteria at 92%. *Gallionellaceae* was unique to this location and present in four of the six samples.

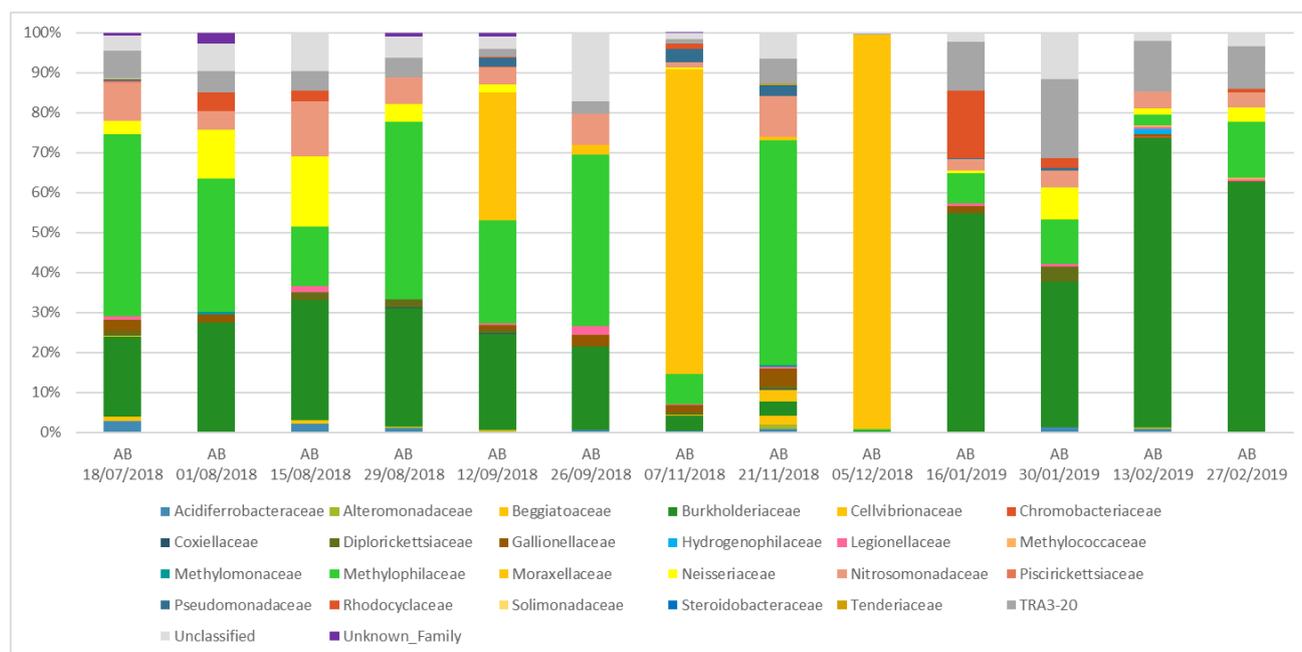


Figure 2.32: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location AgricBuild over the sampling period.

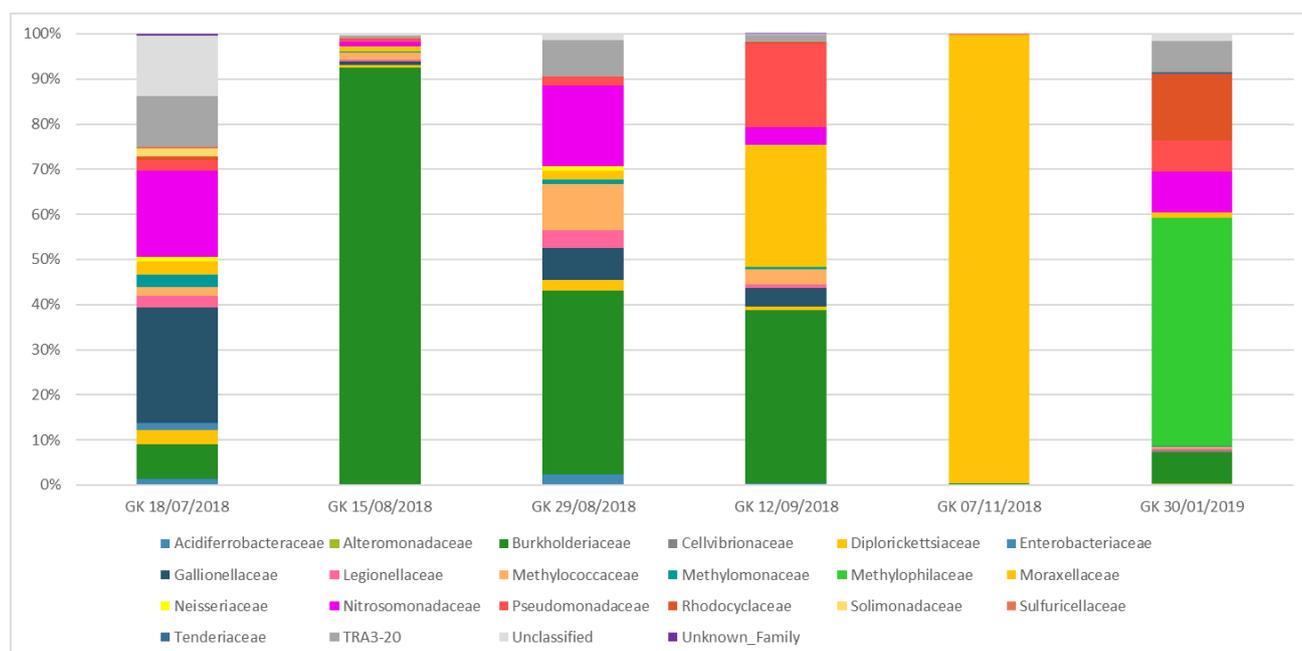


Figure 2.33: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Groenkloof over the sampling period.

The NS2 (Figure 2.34) sample location, which was one of the three samples dominated by the Alphaproteobacteria and the *Sphingomonadaceae* was the most dominant family. *Rhizobiales Incertae Sedis* and *Hyphomicrobiaceae* were also present and shared amongst all samples. Each sample had a unique diversity based on the abundance of families and *Beijerinckiaceae* was present in five of the six samples.

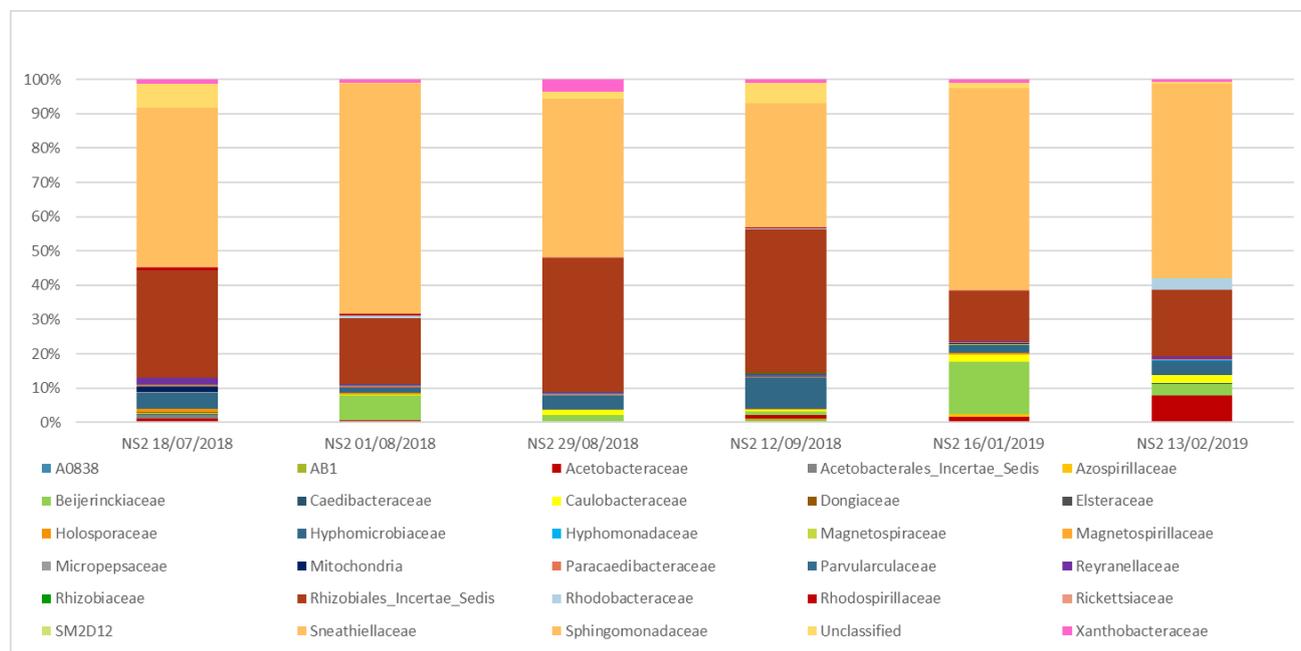


Figure 2.34: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location NS2 over the sampling period.

Silverton (Figure 2.35), like AgricBuild and Groenkloof in this in this group of reticulation sample locations had an abundance of the Gammaproteobacteria. Samples from this location shared *Burkholderiaceae* and *Nitrosomonadaceae*. *Moraxellaceae* dominated a few of the samples and dominated five of the samples taken during the warmer months especially November and February. Most of the samples had a unique community consisting of a range of families.

At Valhalla, similarly to NS2 and Waverley the community was dominated by the Alphaproteobacteria (Figure 2.36). The families shared amongst all the samples were *Hyphomicrobiaceae*, *Rhizobiales Incertae Sedis*, *Rhodobacteraceae* and *Sphingomonadaceae*. For the sample taken in February a relative abundance of 90% for the *Sphingomonadaceae* was observed.

The last sample location, Waverley (Figure 2.37) had a similar distribution to the other samples where each point within the sample had a unique relative abundance of families. The families common at all points were *Beijerinckiaceae*; *Hyphomicrobiaceae*; *Rhizobiales Incertae Sedis* and *Sphingomonadaceae*. *Xanthobacteraceae*, *Acetobacteraceae* and *Caulobacteraceae* were also unique to some samples.

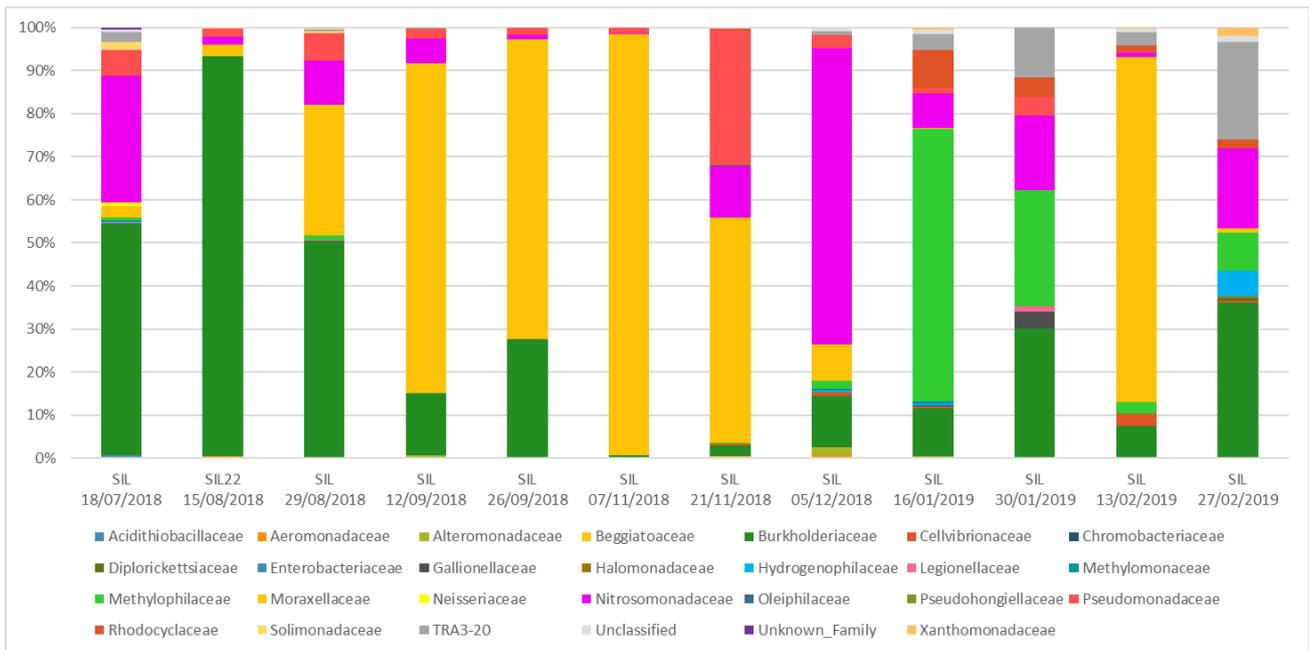


Figure 2.35: Relative abundance plot for the families in class Gammaproteobacteria in phylum Proteobacteria for sample location Silverton over the sampling period.

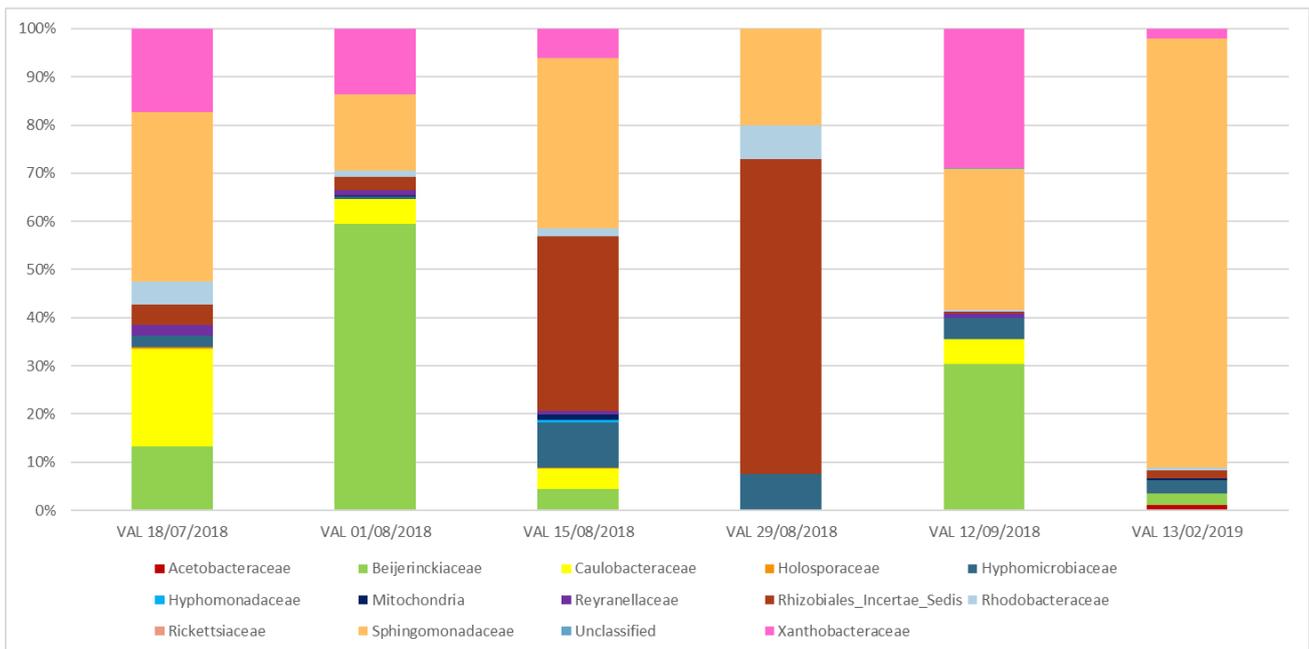


Figure 2.36: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Valhalla over the sampling period.

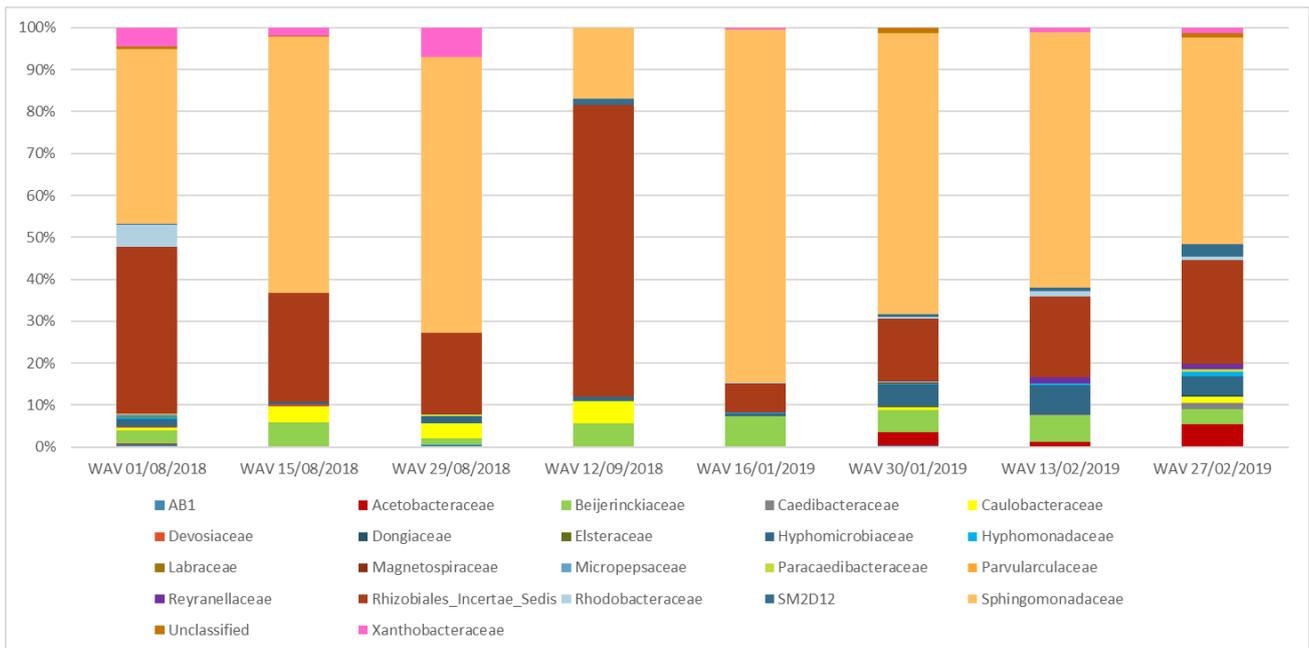


Figure 2.37: Relative abundance plot for the families in class Alphaproteobacteria in phylum Proteobacteria for sample location Waverley over the sampling period.

2.3.2 OTU diversity and distribution

The original dataset consisted of 2657 OTUs and 317572 sequences. The OTUs were taxonomically associated with 9 Phyla, 27 families and 46 genera. There were 88 abundant OTUs found across all 89 samples investigated with 14 of the 88 OTUs shared across all sample locations. Table C1 represents the taxonomic classification of the most abundant OTUs. The most abundant phylum with 86.6% of the OTU sequence abundances was the Proteobacteria. The five most abundant families amongst these abundant OTUs were: *Moraxellaceae* (21.1%); *Burkholderiaceae* (18%); *Sphingomonadaceae* (14.6%); *Rhizobiales Incertae Sedis* (8.2%) and *Pseudomonadaceae* (8.1%) (Table C2). Of the original 2657 OTUs, the dominant OTUs were OTU 2, 6, 3, 5, 8, 17, 19, 20, 1 and 12 which were present at more than 50% of the sampling locations (Table C3). The OTU with the most abundance across all sampling points was OTU 2 which was taxonomically classified as part of the family *Rhizobiales_Incertae_Sedis* and genus *Phreatobacter*.

Spatial and temporal trends across all sampling locations are represented by the Heatmaps in Figure 2.38, 2.39 and Figure 2.40 which were constructed based on MultiCoLA datasets where 1% of the OTUs with the highest abundance in the database was used. The 99% of low-abundance OTUs were removed from the dataset. The Heatmaps were constructed for each of the two types of sample locations to observe the trends among these specific sample locations.

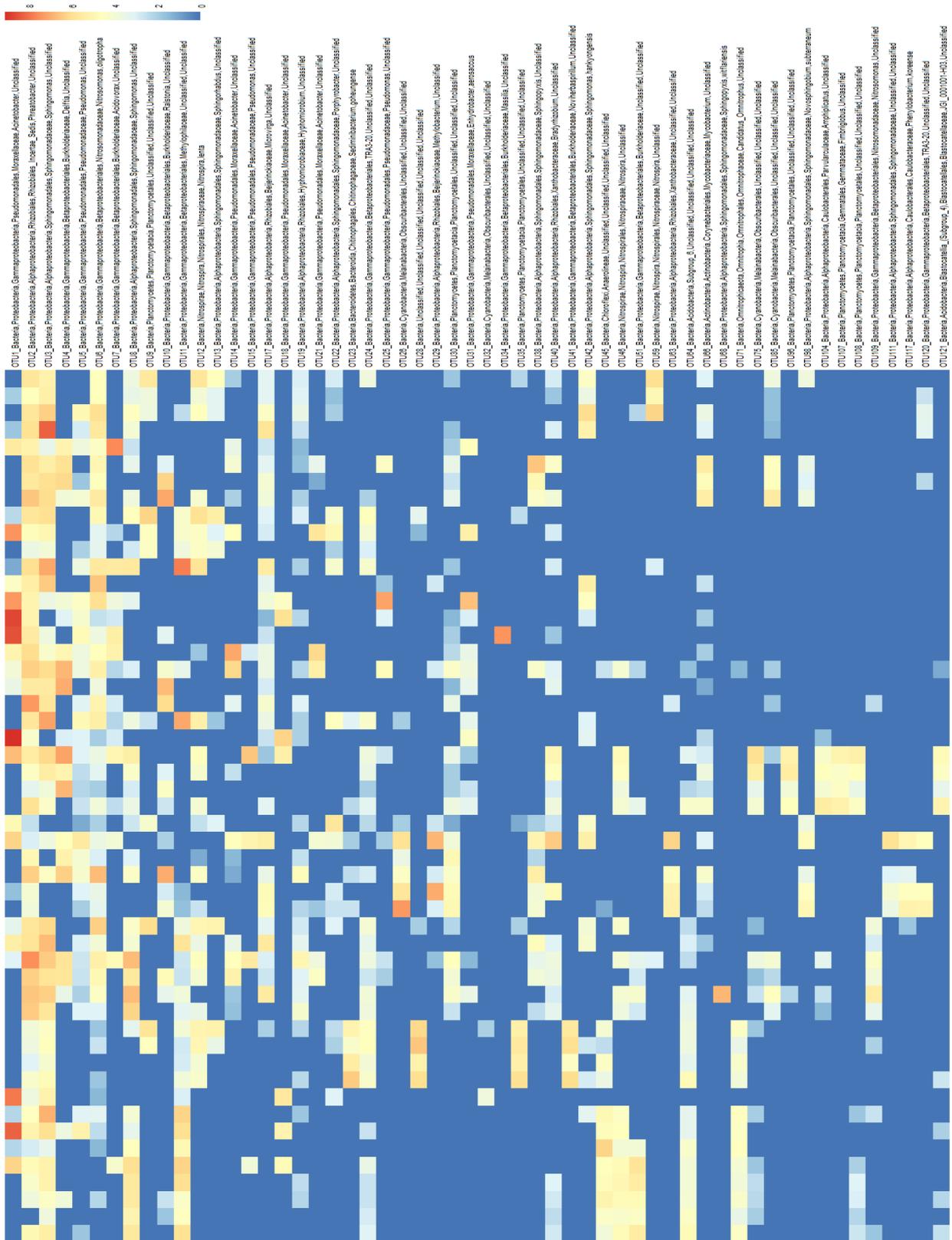


Figure 2.39: Heatmap representing Site 2 samples constructed using the absolute sequence abundances (log transformed) of 87 OTUs across all sampling locations which were selected based on their abundance within a sample (i.e. relative abundance threshold $\geq 1\%$). The Heatmap boxes were coloured from red-to-blue to represent higher and lower abundances.

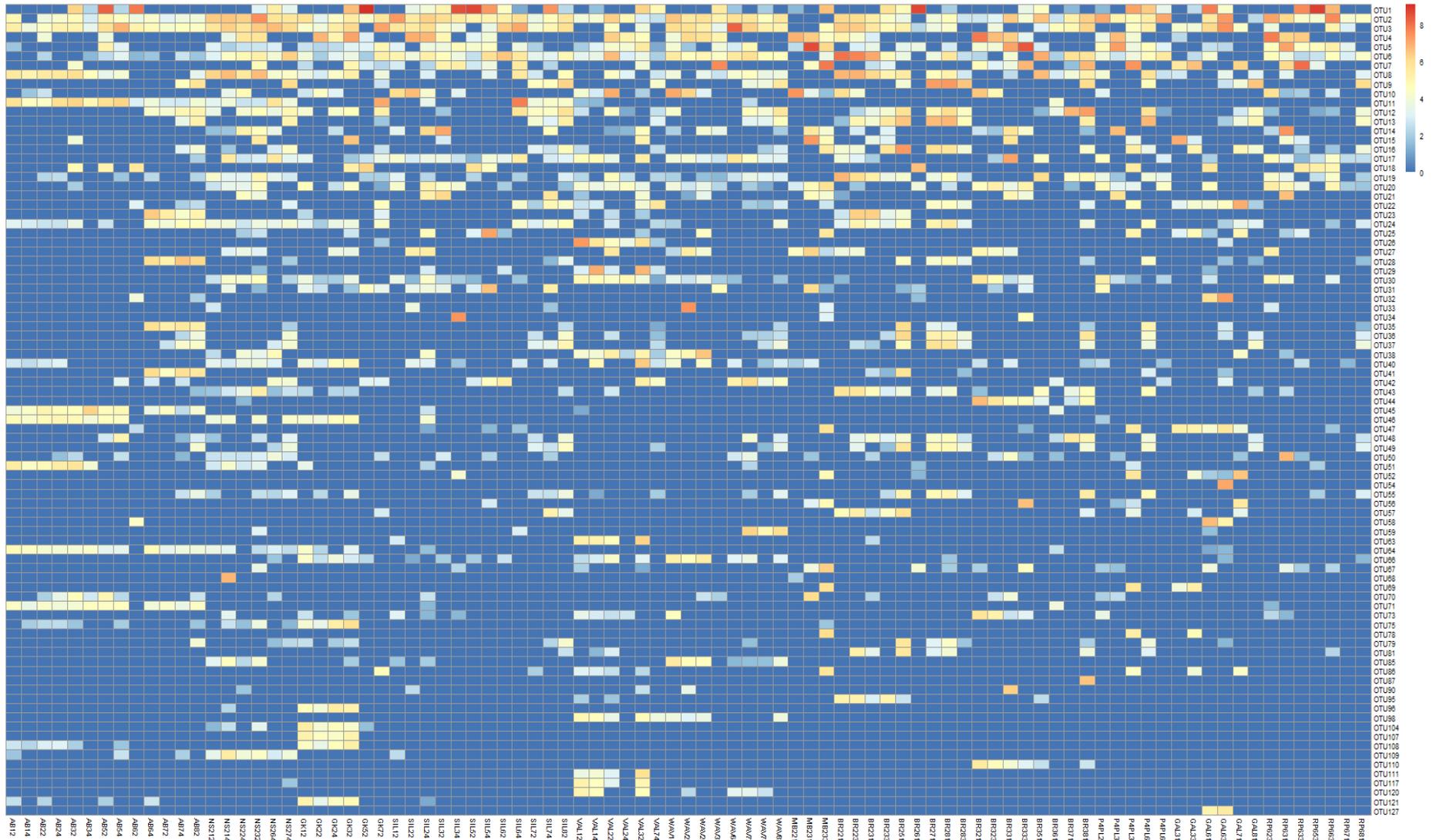


Figure 2.40: Heatmap constructed using the absolute sequence abundances (log transformed) of 87 OTUs across all sampling locations which were selected based on their abundance within a sample (i.e. relative abundance threshold $\geq 1\%$). The heatmap boxes were coloured from red-to-blue to represent higher and lower abundances.

Site 1 (large distribution network) samples (Figure 2.38) were ordered according to the way in which the samples would typically flow in a system, from the treatment to points along the distribution line. A total of 54 OTUs were present in high abundance in the dataset and the family *Burkholderiaceae* was most dominant. There was a few OTUs that were present in all samples, the most abundant of these were OTU 1; 3; 4; 5; 6 and 7. These OTUs were classified as *Acinetobacter*, *Phreatobacter*, *Sphingomonas*, *Delftia*, *Pseudomonas*, *Nitrosomonas* and *Acidovorax* respectively at the genus level. Sample location Ben-Res2-IN (BR2), Brak-RS3-IN (BR3) and Ga-Luka (GL) each had unique clusters of OTUs that were not present in other samples.

The Reticulation system (Figure 2.39) locations had 56 OTUs in the dataset and the *Sphingomonadaceae* was the most abundant family. The most abundant of the OTUs present in all the locations were the same OTUs as the Site 1 sampling locations, with the exception of OTU 2, which was abundant in this dataset. OTU 2 was classified as belonging to the genus *Phreatobacter*. Unique clusters of OTUs were present at location but a few individual samples at specific locations were different in OTU distribution. Silverton (SIL) had more of the dominant OTUs that were highly abundant and had fewer of the less abundant OTUs.

FCM data on these sites (Volume I report) indicated that there were potentially 31 samples where the FCM measurement deviated from the base-line values as measured at the specific sampling location. They represented 9 of the 12 sampling locations, were observed for both distribution (Site 1) and reticulation (Site 2) points and were more common during spring and summer months. The community composition, based on 16S profiling was available for 23 of these samples.

The OTUs commonly present in the distribution samples with higher FCM values included OTU 2, 3, 6, 8, 12, 13, 16, 19, 24, and 43. In the reticulation samples it was OTU 2, 3, 8, 11 and 42. Although they were commonly associated with samples with higher FCM counts, they were not dominant in all those samples and were also present in samples with low FCM measurements. The three OTUs shared between distribution (Site 1) and reticulation (Site 2) systems were OTU 2 (*Phreatobacter* sp.), OTU 3 and OTU 8 (both *Sphingomonas* spp.).

2.3.3 Alpha diversity

The alpha diversity which represents the diversity at each sample location showed no real seasonal trends over the sampling period. The richness of the OTUs were not influenced by the seasons or temperature directly.

The Observed, Shannon, Simpson and ChaO1 diversity measures were investigated (Figure 2.41) at each location. Based on the Observed OTUs, which takes the unique OTUs into account, AgricBuild (AB), Groenkloof (GK) and NS2 had a notably higher diversity than the other sampling locations. AgricBuild had the highest observed diversity numbers. The ChaO1 diversity measure which takes the abundance into account, confirmed that these three sampling locations previously mentioned had higher richness and diversity. These locations also had rarer OTUs in their communities.

The sample Brak-RS3-IN (BR3) in the distribution system dataset had the highest diversity amongst the locations for the Shannon measure. The Shannon diversity measure based on the abundance and evenness of samples showed that sampling locations AgricBuild (AB), Groenkloof (GK), NS2 and Silverton (SIL) had high values (>3) which indicated a higher diversity and greater evenness within those samples due to the presence of rare OTUs at these locations. The Simpson diversity measure that represents the number of species and their relative abundances showed the increase of diversity as the evenness increases. The majority of the sampling locations had a high diversity. The locations that were outliers at Site 1 dataset was Ben-Res2-IN (BR2) and Rust-P6 (RP6). The Reticulation points that were outliers were namely AgricBuild (AB), Groenkloof (GK) and Silverton (SIL). All the outlier points were present during spring.

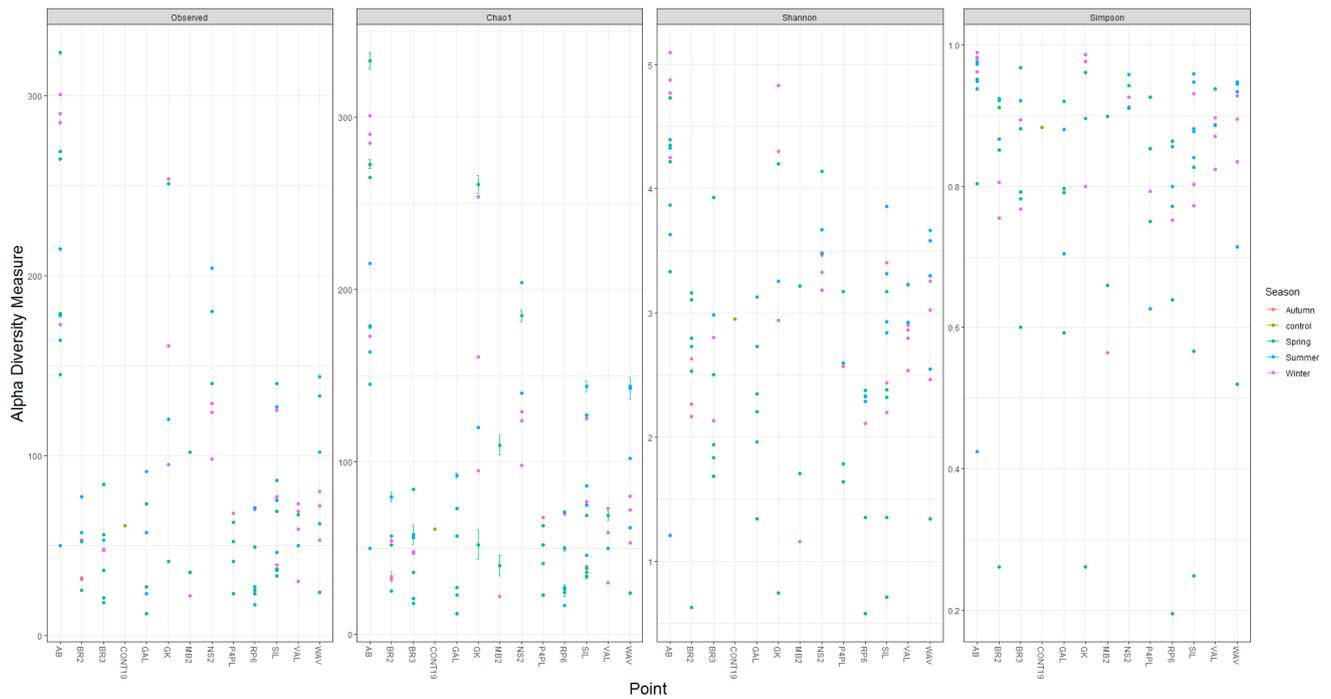


Figure 2.41: Alpha diversity represented through Observed species, Chao1, Shannon and Simpson values for all samples taken at the 12 sampling points.

2.3.4 Beta diversity

The beta diversity which shows the community structure and the diversity between sample locations is represented in Figures 2.42 and 2.43.

The beta-diversity was calculated using the Fast Parallel UniFrac (Hamady *et al.*, 2010) algorithm in Phyloseq and distances based on a phylogenetic tree were generated in Ape (Version 5.3) using the OTU sequences. The weighted Unifrac (Figure 2.42) which represents a quantitative analysis of the community structure and considers the absence/presence of taxa and their abundances between the samples showed temporal changes. A clustering of some spring and winter samples were observed at the different locations. The other values were mostly random with some overlap between winter and summer.

The unweighted Unifrac (Figure 2.43) represents the qualitative analysis of the community's membership and shows the pairwise dissimilarity distances between the individual samples from the different sampling locations. This analysis includes the presence/absence of taxa between pairs of samples. A clustering of seasons is observed where spring and winter formed small clusters within the samples but overlap was seen between winter and spring as well as summer samples.

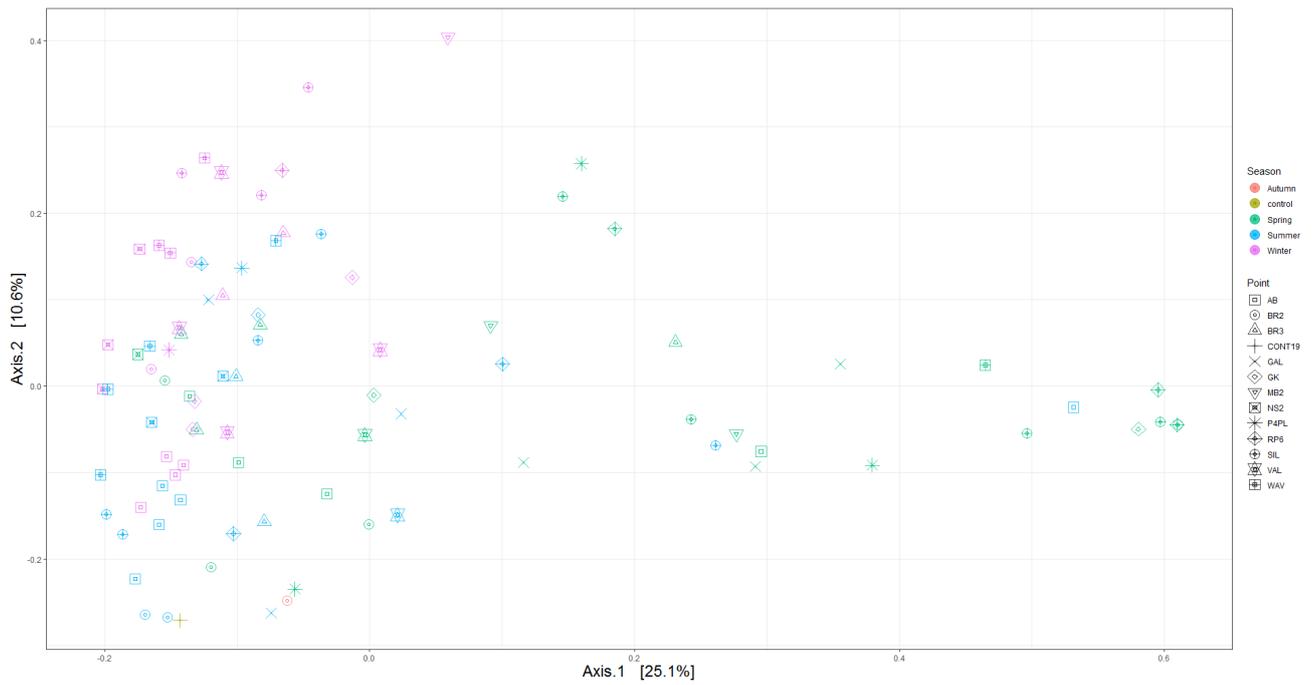


Figure 2.42: MDS/PCoA plot on weighted-UniFrac distance representing the beta diversity of the community relationship between 12 sampling locations over an 8-month period.

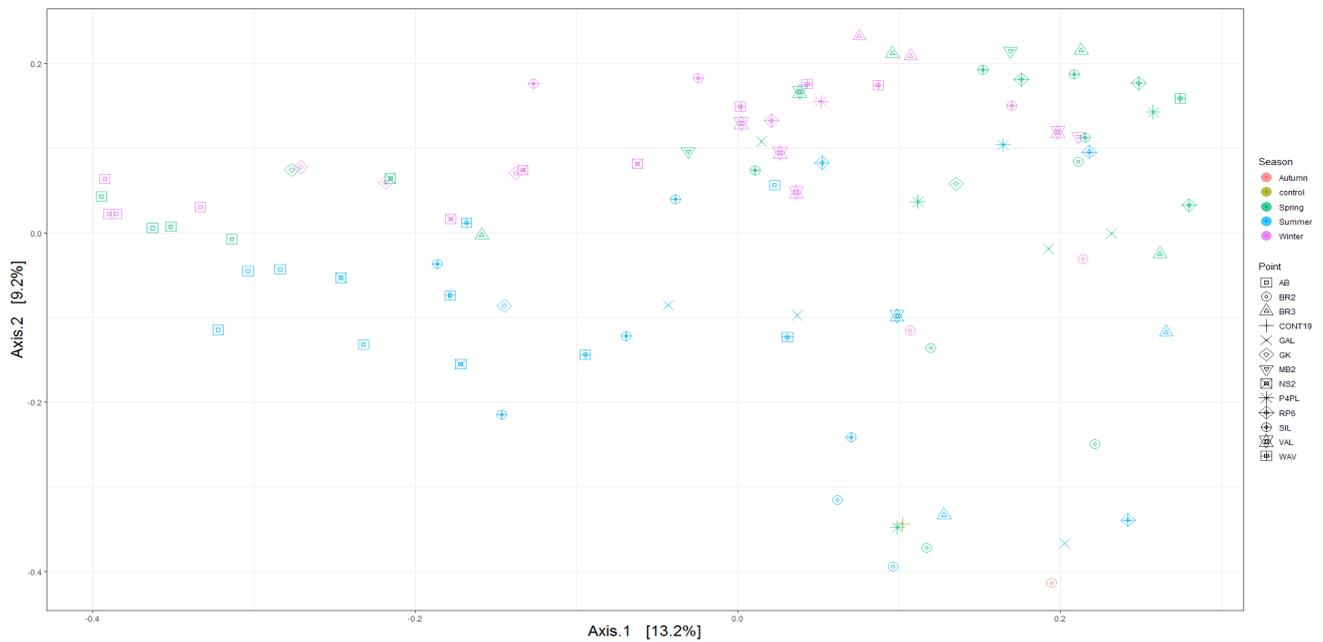


Figure 2.43: MDS/PCoA plot on unweighted-UniFrac distance representing the beta diversity of the community relationship between 12 sampling locations over an 8-month period.

2.4 BACTERIAL DIVERSITY IN DRINKING WATER SAMPLES BASED ON ISOLATES OBTAINED FROM HETEROTROPHIC PLATE COUNT ANALYSES

Single colonies were picked and purified from the heterotrophic plates. Once pure cultures were obtained, isolates were identified through Sanger sequencing of the 16S rRNA region.

In total, 372 isolates were successfully purified, and 368 isolates were successfully sequenced with sequences that were of good quality to perform the downstream analysis. Of the 368 isolates 277 (75%) were originally from R2A plates. A diverse bacterial community was observed and the isolates belonged to 53 genera and 28 families of which the majority are known to be found in drinking water.

The 11 families which contained singletons were *Beijerinckiaceae*, *Chitinophagaceae*, *Enterobacteriaceae*, *Flavobacteriaceae*, *Hyphomicrobiaceae*, *Microbacteriaceae*, *Nocardiaceae*, *Rhizobiaceae*, *Rhodobacteraceae*, *Rhodocyclaceae* and *Rhodospirillaceae*. The remaining isolates were part of 17 families, e.g. *Acetobacteraceae*, *Bacillaceae*, *Bradyrhizobiaceae*, *Caulobacteraceae*, *Comamonadaceae*, *Erythrobacteraceae*, *Intrasporangiaceae*, *Lysobacteraceae*, *Methylobacteriaceae*, *Micrococcaceae*, *Moraxellaceae*, *Mycobacteriaceae*, *Oxalobacteraceae*, *Phyllobacteriaceae*, *Pseudomonadaceae*, *Sphingomonadaceae* and *Staphylococcaceae*. These isolates were identified using phylogenetic analyses.

The four most prominent families will be discussed below.

The family *Sphingomonadaceae* had the most isolates with 117 (32%) associated with the family. Nearly half of these isolates (57) were associated with strains from the genus *Sphingomonas* and were isolated 10 of the 12 sampling points. Fifteen of the isolates from the sample point AB grouped with *Sphingomonas ginsenosidimutans*. These isolates also included members belonging to *Sphingomonas aquatica*, *Sphingomonas aerolata*, *Sphingomonas faeni*, *Sphingomonas aurantiaca* and the potential pathogen *Sphingomonas ursincola*. Forty of the strains belonged to the genus *Sphingopyxis* and were isolated from 8 of the sampling points. They were identified as *Sphingopyxis ginsengisoli*, *Sphingopyxis taejonensis*, *Sphingopyxis alaskensis*, *Sphingopyxis indica*, *Sphingopyxis flava* and *Sphingopyxis panaciterrae*. Seventeen isolates grouped with the type strains of the genus *Sphingobium*, which is known to be important environmental isolates that play a role in biodegradation and bioremediation of pollutants (Young *et al.*, 2007). The type strains they grouped with were *Sphingobium hydrophobicum*, *Sphingobium xenophagum*, *Sphingobium olei* and *Sphingobium naphthae*. Fourteen of the isolates originated from the household in Waverley. The last genus that formed part of the *Sphingomonadaceae* family was *Novosphingobium* with three isolates obtained from the distribution system.

The *Methylobacteriaceae* was the second most common family with 52 isolates (14%) belonging to the genera *Methylobacterium* and *Methylorubrum*. Most of *Methylobacterium* strains grouped closely several of the type strains and were identified based on a blast search as *Methylobacterium bullatum*. Six of the isolates grouped with three potential pathogens namely *Methylobacterium fujisawaense*, *Methylobacterium mesophilicum* and *Methylobacterium radiotolerans*. Five of these isolates were isolated from the reticulation system at Valhalla. The *Methylorubrum* isolates (14) originated from four of the sampling points with 11 of the isolates obtained from sampling point AB.

Forty-two isolates (11%) belonged to the family *Mycobacteriaceae*. Some of the isolates grouped closely with five potentially pathogenic *Mycobacterium* type strains. These 5 pathogenic strains were *Mycobacterium abscessus*, *Mycobacterium chelonae*, *Mycobacterium cosmeticum*, *Mycobacterium frederiksbergense* and *Mycobacterium mucogenicum*.

The family *Comamonadaceae* was represented by 27 isolates (7%). Twenty-one of the isolates from different sampling points did not group close to any type species which may imply that these isolates could be represent new species and need to be further investigated. The type strains of the species *Acidovorax temperans*, *Acidovorax radialis* and *Acidovorax defluvii* grouped closely with 5 isolates from sample points Ga-Luka, SIL, P4PL and BR3. A few isolates grouped with the type strains of the species *Brevundimonas aurantiaca*, *Brevundimonas intermedia*, *Brevundimonas nasdae*, *Caulobacter segnis* and *Caulobacter vibrioides*.

Overall, the most common bacteria obtained from the twelve locations grouped with the families *Sphingomonadaceae*, *Methylobacteriaceae*, *Mycobacteriaceae* and *Comamonadaceae*. The diversity of the families varied across the sampling locations and some families had more isolates from specific points. Together they represented 2/3 of the total number of isolates obtained during this study.

2.5 SUMMARY

2.5.1 Relative abundances of OTUs making up bacterial communities in the distribution systems

The bacterial community for the selected sampling locations in this study were diverse. From the relative abundances and the Heatmaps, the most dominant phyla which was shared amongst all twelve sample locations was the Proteobacteria. The dominance of Proteobacteria was consistent with literature (Bautista-de los Santos *et al.*, 2016; Lu *et al.*, 2013). The classes Alpha- and Betaproteobacteria were shown to be dominant in drinking water distribution system studies. In this particular dataset, the classes Alpha- and Gammaproteobacteria were dominant. This difference is only due to the fact that in the latest taxonomic scheme the "Betaproteobacteriales (formerly known as Betaproteobacteria) is now Burkholderiales, an order of Gammaproteobacteria" (Parks *et al.*, 2018). The sequencing within this chlorinated system includes the dead bacteria which may have completely overshadowed what is really happening. Viability-sequencing methods are however quite tricky to use and were not used in this study because the complete microbial community was investigated.

The dominant organisms are known to be determined by disinfectant regimes; temporal changes and multiple other factors (Revetta *et al.*, 2010; Holinger *et al.*, 2014; Prest *et al.*, 2014; Williams *et al.*, 2004). The variation and complexity in the bacterial community are a common trend observed. Berry *et al.* (2006) stated that "Alphaproteobacteria were the dominant isolates in both chloraminated and chlorinated water from model DSs, whereas Betaproteobacteria were found to be more abundant in chloraminated water than in chlorinated water". Five of the six distribution system locations in this study showed Gammaproteobacteria and the order Betaproteobacteriales as more abundant. These locations were linked to free chlorine residual concentrations of around 0.4 mg/L.

The samples shared common families and there was no specific trend for samples at each location. Each sample set had unique distribution of the families amongst the sampling points. Site 1 dataset was generally more abundant in Gammaproteobacteria, with the order Betaproteobacteriales consisting of *Burkholderiaceae* as the most dominant in the dataset. The Reticulation system had an even spread of the Alpha and Gammaproteobacteria classes.

The variation within the community at each sampling location was an interesting trend. The relative abundance at the family level showed that each location had diversity within the samples but that some sampling locations shared the same top three dominant families. Sampling locations NS2; Valhalla and Waverley had *Sphingomonadaceae*, *Rhizobiales_Incertae_Sedis* and *Beijerinckiaceae* in common. These

families were present at sampling locations which represented the point of use or reticulation locations and could imply that pipelines feeding these areas could possibly be connected or that they were exposed to similar abiotic conditions. Members of the *Sphingomonadaceae* family made up 14.6% of the families observed in the OTUs taxonomic classification and are commonly observed in drinking water systems but additionally may be observed in various other environments (Luhrig, 2016; Hwang *et al.*, 2012; Vaz-Moreira *et al.*, 2013; Ling *et al.*, 2016). These bacteria are also known to colonize drinking water taps, be resistant to chloramine and chlorine, adaptable to varying temperatures and may also be initiators for biofilm formation (Luhrig, 2016; Vaz-Moreira *et al.*, 2013). *Rhizobiales_Incertae_Sedis* are known to be denitrifying bacteria and exist in anaerobic environments and waters with a depth more than 30cm (Jia *et al.*, 2019; Yang *et al.*, 2019). Taxa which form part of the family *Beijerinckiaceae* are known to be methanotrophic and prevail in acidic habitats as well as wetlands emitting methane (Ulrich *et al.*, 2018; Tamas *et al.*, 2014).

Sampling points Groenkloof and Silverton and P4PL had the families *Burkholderiaceae*, *Moraxellaceae* and *Nitrosomonadaceae* in common. *Burkholderiaceae* isolates were detected in source-to-distribution drinking water in Australia and Italy and an increase in abundance was observed in biofilms and chlorinated water, the bacteria are also stated as being resilient growers (Zhang, 2012; Tsao *et al.*, 2019; Kaestli *et al.*, 2019; Bruno *et al.*, 2018). Bacteria from the family *Moraxellaceae* have been observed in tap water and groundwater samples (Zhang, 2012; Bifulco *et al.*, 1989). Certain members of the genus *Acinetobacter* which forms part of this family are known to be potential pathogens and sewer indicator bacteria detected in water purification systems (Yang *et al.*, 2019; Penna *et al.*, 2002; Zhang, 2012). Members of the *Nitrosomonadaceae* family are ammonia metabolisers and detected in chloraminated systems because of the presence of ammonia present as a result of chloramination (Holinger *et al.*, 2014). These bacteria are known to be present in source waters but can withstand treatment and thrive within the distribution network (Hwang *et al.*, 2012; Wang *et al.*, 2015).

Other families that were dominant amongst sampling locations were *Pseudomonadaceae*, *Methylophilaceae*, *Hyphomicrobiaceae* and *Rhodobacteraceae*. *Pseudomonadaceae* was shown to be more dominant at lower temperature (22°C) chloraminated water systems and could consist of pathogens like *Pseudomonas aeruginosa* (Hwang *et al.*, 2012; Gensberger *et al.*, 2015; Shar *et al.*, 2011; WHO, 2011). *Methylophilaceae* bacteria are commonly found in the environment and have the ability to utilize organic compound including Cl-compounds and refractory contaminants (Rozej *et al.*, 2015; Buse *et al.*, 2014; Hwang *et al.*, 2012). Some taxa associated with *Hyphomicrobiaceae* are known for the removal of nitrogen and shown to be part of biofilm development in drinking water (Jia *et al.*, 2019; Zhang, 2012; Chao *et al.*, 2015). *Rhodobacteraceae* can exist in stagnant tap water or water treatment plants and are amoeba-resisting bacteria found to be more abundant in biofilms made of Cu piping (Zhang, 2012; Lu *et al.*, 2014; Yang *et al.*, 2019).

The top three most abundant OTUs in the dataset were OTU 2 classified on a genus level in *Phreatobacter*, OTU 6 classified as *Nitrosomonas oligotropha* and OTU 3 classified to the genus level as *Sphingomonas*. The most abundant OTU taxonomically classified as *Phreatobacter* was shown to be isolated from ultrapure water of purification systems and the type strain *Phreatobacter oligotrophus* was described in 2014 (Toth *et al.*, 2014). This OTU had a 91% occurrence in the dataset across all the sampling locations. The other abundant OTUs are also microbes generally found in drinking water and some members could potentially be opportunistic pathogens. The occurrence of these OTUs are determined by the treatment practices and the conditions in the distribution and reticulation network.

Generally, no significant changes were observed in the alpha diversity analysis and the microbial communities at different sample locations grouped at random across the sampling date and location. The diversity at each sampling point was however very high at some sampling locations, influenced by the presence of rare OTUs or low diversity and species richness at other locations. The microbial communities across the sampling locations from the same sampling sets (Site 1 or 2) were not consistent with each other and most locations displayed a unique trend.

The beta diversity results showed that there were seasonal groupings between some samples taken at different locations. The clustering included winter, summer and spring which were the seasons during which the sampling occurred. Most of the samples overlapped over the different seasons and a weak correlation was seen with the weighted and unweighted Unifrac plots where the principle coordinate axes values of the samples were less than 1. For the unweighted Unifrac analysis, a few points clustered together showing similarity between these sampling locations. The sampling locations that clustered together were Valhalla, Waverley, Groenkloof, Silverton and production point MB2. Four of these points were from the point of use sample set. This could be expected as these locations represented tap samples from households. For the weighted Unifrac, the spring and summer sampling seasons did not form a close group but a large majority of the samples clustered together. Samples from the distribution system and points of use formed groups and the winter sampling season also clustered together indicating a similarity in the communities. The overlapping of seasons and sampling points was the trend observed between the sampling locations.

The microbial communities were not influenced seasonally or temporally, rather the system treatment, source water, storage, pipe condition, maintenance and other factors may have all played a role in influencing the water quality and consequently the microbial community present. The use of different disinfectants and treatments may have varied at different sampling locations and consequently influenced the community that occurs at each location. The two locations collected from the AgricBuild and NS2 on the University of Pretoria Hatfield campus were supplied with water from a catchment tank and this may have shaped the community through possible stagnation in the tanks and contamination, should the tank have been exposed.

2.5.2 Bacterial diversity in drinking water samples based on isolates obtained from heterotrophic plate counts

In this study, a culture-dependent approach for the identification of pure colonies from heterotrophic plate counts was carried out. The bacterial community's diversity from a phylogenetic perspective was determined based on these isolates. The media used to grow the single colonies played an important role in the types of bacteria that were identified. A few families had isolates which only grew on one of the media. R2A media favoured the growth of *Bradyrhizobiaceae*; *Comamonadaceae*; *Phyllobacteriaceae*; *Pseudomonadaceae* and *Oxalobacteriaceae*. YEA favoured the growth of *Bacillaceae* and *Staphylococcaceae*. For 22 families, isolates were obtained from both media. The sampling locations had diverse communities and few families favoured specific locations.

Sanger sequencing was used to obtain the forward region of the isolate's 16S rRNA sequence. The small size of the sequences obtained (~300 bp) for the forward region was one of the reasons some of the identifications could only be done up to the genus level. The fact that the V4 variable region across all bacteria is less variable than some of the other regions could be another reason for the lack of resolution in the trees.

Sphingomonadaceae was the most dominant family overall and had a large number of isolates from 10 of the 12 sampling points. The *Sphingomonas* bacteria were most commonly isolated from the reticulation system which has smaller, narrower pipes which favours the proliferation of biofilms. *Sphingobium hydrophobicum* is a hydrophobic bacterium which was isolated from electronic-waste recycling sediment in China and is closely related to other species of the genus *Sphingobium* (Chen *et al.*, 2016). *Sphingobium xenophagum* was described as an "eater" of xenobiotic compounds (Pal *et al.*, 2006). *Sphingobium olei* was isolated from oil-contaminated soil in Taiwan and cultured on nutrient agar (Young *et al.*, 2007). *Sphingobium naphthae* has the ability to degrade aliphatic hydrocarbons and is closely related to *Sphingobium olei* and was isolated from a diverse range of habitats including dump sites, river sediments and copper mine soil to name a few (Chaudhary *et al.*, 2017). The potential pathogen *Sphingomonas ursincola* was isolated from drinking water from taps. *Sphingomonas aerolata*; *Sphingomonas faeni* and *Sphingomonas aurantiaca* are able to survive in harsh environments and were isolated from hydrocarbon-contaminated soils and whirled-up dust in a cow

barn and cultured on R2A plates; which 2 SIL isolates that grouped with these species also cultured on (Busse *et al.*, 2003). *Sphingomonas aquatica* was isolated from tap water in South Korea and one isolate from sample point GK cultured on R2A grouped with this strain (Choi *et al.*, 2017). *Sphingomonas ginsenosidimutans* strain which was initially described by Choi *et al.* (2017) was amended and the strain tested positive for the hydrolysis of aesculin which is a glycoside that occurs in plants (Feng *et al.*, 2018). Strains in *Sphingopyxis* genus are known to occur in habitats like soil and underground water (Takeuchi *et al.*, 2001). *Sphingopyxis alaskensis* grouped with an isolate from sample point SIL and was isolated from a river's subsurface polluted with chlorophenolic compounds in Chile, the strain was cultured on R2A but the isolate grouping with this strain was cultured on YEA which may suggest it can grow on different media (Godoy *et al.*, 2003). *Sphingopyxis panaciterrae* grouped with isolates from sample point Wav and NS2 and was isolated from soil in a ginseng field in South Korea (Lee *et al.*, 2008). *Novosphingobium fuchskuhlense* was isolated from a basin/ subsurface water of a lake in Germany and forms capsules.

Methylobacterium was described in 1976 by Patt *et al.* as a species of methane-oxidizing bacteria which can also utilize complex organic carbon (Patt *et al.*, 1976). These bacteria may have been present in the sample as a result of contamination from the environment. The description of this species was said to be based on *Methylobacterium organophilum* and habitats such as soil, freshwater, lake sediments and hospital environments are where they are isolated from (Green and Bousfield, 1983). Isolates from sample point Val and RP6 grouped separately with *Methylobacterium isbiliense* which was isolated from drinking water in Spain and described by Gallego *et al.* (2005). All 3 of the *Methylobacterium* pathogens have been associated with drinking water from taps and are found in biofilms as well as immunocompromised patients with nosocomial infections (Green and Bousfield, 1983).

The *Mycobacteriaceae* isolates were found in many previous studies and investigated for their infectious abilities. The genus *Mycobacterium* was recently reclassified and a number of new genera was proposed. *Mycobacterium abscessus* was divided into subspecies (Tortoli *et al.*, 2016); *Mycobacterium chelonae* was renamed as *Mycobacteroides chelonae*; *Mycobacterium cosmeticum* was renamed to *Mycolicibacterium cosmeticum* (Cooksey *et al.*, 2004); *Mycobacterium frederiksbergense* was renamed to *Mycolicibacterium frederiksbergense* and *Mycobacterium mucogenicum* was renamed to *Mycolicibacterium mucogenicum*. *Mycobacterium abscessus* was isolated from sputum cultures (Tortoli *et al.*, 2016). *Mycobacterium chelonae* was first isolated from a turtle in 1903 by Freidmann. *Mycobacterium cosmeticum* was isolated from "a granulomatous subdermal lesion of a female patient in Venezuela who was undergoing mesotherapy with an unknown substance(s) for a cosmetic purpose (weight loss)" (Cooksey *et al.*, 2004). *Mycobacterium frederiksbergense* was isolated from coal tar contaminated soil and has the ability to degrade polycyclic aromatic hydrocarbons. *Mycobacterium mucogenicum* was isolated from patients' abdominal cavity, automated dialysis machines, and tap water used to supply the machines.

Based on the data provided in Section 2.3, three OTUs were commonly shared between the distribution and reticulation systems. They were OTU 2 (*Phreatobacter* sp.), OTU 3 and OTU 8 (both *Sphingomonas* spp.). Isolates belonging the genus *Sphingomonas* were also the strains most commonly isolated from the sampling points during this study. *Sphingomonas* species are widely distributed in the environment and around 130 species have so far been formally described. They are well known for their ability to degrade various aromatic hydrocarbons (Asaf *et al.*, 2020). Their specific role in drinking water distribution systems will have to be further investigated. It was not surprising that no *Phreatobacter* strains were isolated as it is well known that the isolation of oligotrophic bacteria is difficult and that many of them have not yet been cultivated. This genus was first isolated for ultrapure water containing limited amounts of organic and inorganic compounds (Toth *et al.*, 2014) and finding it in drinking water systems at a high relative abundance is not surprising.

CHAPTER 3: IMPACT OF THE RESERVOIR DESIGN AND MANAGEMENT ON THE MICROBIAL COMMUNITY

3.1 INTRODUCTION

Globally there is a need for safe drinking water. The demand for drinking water has increased due to a rapidly growing population. According to the Water Organisation, 844 million people lack access to safe drinkable water and addressing this issue has been a sustainable development goal (Bain *et al.*, 2012). The lack of safe water leads to the increased spread of infections and diseases, causing millions of deaths (Messner *et al.*, 2006). Access to safe treated drinking water would directly improve health and decrease the spreading of diseases. Efforts to improve treatment and distribution should be a high priority for governments, local authorities and communities. To ensure a safe and clean water supply, monitoring, evaluating, and quality control checks are of great importance. Drinking water quality can be determined by assessing the chemical, physical and biological composition of the water (Craun and Calderon, 2001). Water distributors often examine the chemical characteristics but monitoring the biological composition can be more difficult (Craun and McCabe, 1973).

Although the drinking water system has been established for many years and has supplied safe drinking water to consumers, there is a growing demand on supply due to the increasing population. Therefore, understanding the drinking water design and operations would aid in ensuring safe drinking water is continuously supplied by illustrating any potential microbial risks present in the system and allowing for recommendations to improve the drinking water systems design while evaluating a methodology for testing microbial quality. The World Health Organization has proposed that securing the microbial safety of water supplied to communities should be based on multiple barriers to prevent or limit contamination or regrowth (Edokpayi *et al.*, 2018, Bain *et al.*, 2014, Prest *et al.*, 2016a). Barriers include protection of the source, treatment, disinfection and management of distribution and supply networks. Although treatment and disinfection are critical barriers, maintaining the quality during distribution is also crucial; this can prove to be challenging, as various factors can change the quality, especially the microbial quality.

Studying the microbial ecology of these systems will allow for a better understanding of the microbial interactions within the community and their environment and how drinking water systems could be best managed to supply safe water to consumers (Beszteri *et al.*, 2010). However, knowledge of the community members and their relative abundance is not enough. It is also essential to know what metabolic activities and processes these communities are involved in and the functional contribution of individual community members to these processes. Currently, metagenomic and transcriptomic studies are used to determine the function and behaviour of microorganisms in drinking water systems (Zhou *et al.*, 2015). Systems in our distribution line supply water to consumers that are hundreds of kilometres away from the source water. This water would need to be stored and slowly pumped into different community areas. Water is typically stored in large cementitious tanks referred to as service or community reservoirs and allows for water storage over long periods (Prest *et al.*, 2016b). The reservoirs' geometry and schematics play a role in the quality of the water, as well as a change in that quality.

Community reservoirs can be described as storage reservoirs which keep the balance between supply and demand in a distribution system (Zhang *et al.*, 2014). As service reservoirs store water, the water ages and the quality of the water can become questionable (Fard and Barkdoll, 2018). The reduction in water quality in these reservoirs is becoming an increasing problem for municipalities that are responsible for supplying clean drinking water (Rossman *et al.*, 1995; LeChevallier *et al.*, 1996; Marek *et al.*, 2007; Fard and Barkdoll, 2018).

The approach to design large reservoirs is very conservative and it can consequently reduce the water quality (Basile *et al.*, 2008; van Zyl and Haarhoff, 2007). Because of the large nature of service reservoirs, it is nearly impossible to have a completely mixed flow and mixing energy of the water entering the system is too low to promote optimal mixing for minimal loss of disinfectants (Zhang *et al.*, 2013). Stagnated zones, often referred to as dead zones describes to areas with slow flowing water or water that does not flow at all. These stagnated zones can cause negative public health effects (Fard and Barkdoll, 2018). For the water quality to be acceptable, the number of disinfectants needs to be sufficient. The disinfectants can be consumed by microorganisms, organic impurities, deposits, corrosion products and ammonium and metal compounds when water is not mixed properly and thus water age is increased (Codina *et al.*, 2015, Moncho-Esteve *et al.*, 2015). Therefore, the disinfectant residual does not reach the stagnated zones and microbial growth is enhanced. South Africa's storage water is found to have high levels of microbial contaminants (Luyt *et al.*, 2012). According to Turton (2008), the amount of microcystin (algae that forms when conditions are favourable for microbial growth) in service reservoirs in South Africa is amongst the highest in the world.

Previous research has identified community reservoirs to also impact the overall microbial quality within the system. It is, therefore, vital to develop a better understanding of the operation of these systems and how this impact on the microbial community and activities associated with these reservoirs. It is only through a comprehensive understanding of the ecology within the system and influences from environmental elements in the system that appropriate management practice could be put in place to ensure safe drinking water (Benson *et al.*, 2017).

Mixing in the reservoir is essential to ensure no dead zones are in the reservoir. This would, in turn, lead to reduced microbial growth, as stagnant waters can lead to favourable conditions for microbial growth (Boulous *et al.*, 1996). Shorter and reservoir tanks covering a large surface area are typically less susceptible to mixing than higher, more narrow tanks. Mixing within the reservoir can be optimized by reducing the diameter of the inlet, as this would increase the force of the inflow and maximize volumetric exchange while the reservoir is filling up (Martel *et al.*, 2002). Poor mixing and flow in the reservoir could allow for stagnant waters, leading to low disinfection residuals, high disinfection by-products and nitrification in chloraminated systems (Hack, 1984), creating optimal conditions and nutrients for microbial growth.

Reservoirs should be kept nearly full to allow the drinking water system to respond to the high demand (Edwards and Maher, 2008). The inlet, outlet, hydraulic pressure, flow in the reservoir, positioning, height, width, dead zones, temperature and mixing in the reservoir are all properties that can affect the microbial quality and growth in the reservoirs (Edwards and Maher, 2008). To observe and understand the quality of the water in reservoirs would require sampling from different zones and different depths. A study done by Kennedy *et al.* (2001) showed that taller reservoirs have more dead zones than shorter or narrower reservoirs. Dead zones are problematic in a reservoir as microbial activity and growth occur.

Changes and differences in thermal reading within the reservoir can influence water quality. Increased temperature in the reservoir can allow for increased decay in chlorine residuals, allowing for the regrowth of microbes in the system (Kirmeyer *et al.*, 1995). During the warmer months, i.e. summer and spring, reservoirs do not fill up adequately due to the high demand, leading to inconsistencies and lower residence time in the reservoir, which can lead to the water quality not being the same throughout the year (Mahmood, 2005). Storage for long time frames allows for a stagnant environment, increasing the microbial growth potential. Stagnant waters typically have nutrient content, are warmer, have little to no pressure and minimal disruptions, allowing for growth. Coliform production is specifically found in high abundance due to warmer conditions in reservoirs.

Biofilm formation is known to occur in reservoirs (Chowdhury, 2009). A study done by Lliros *et al.* in 2014 showed that biofilm formation occurred more in stagnant waters than flowing waters. They found that waters

with higher velocities tend to have a lower concentration of bacterial cells due to their ability to cause cell detachment (Lliros *et al.*, 2014). An increase in residence time correlates with increased bacterial abundances due to the 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 Total organic compounds. These factors caused significant shifts in the microbial community.

Depleted chlorine readings can be seen in the upper zones of the reservoir, harming the water quality, as microbial growth could occur. Free chlorine often has a low residual in the reservoir as it gets depleted quickly (Sawyer *et al.*, 2003). The decay has also been reported to deplete faster due to change in pH, temperature, flow velocities and water quality coming in (Digiano and Zhang, 2005; Hallam *et al.*, 2003; Powell *et al.*, 2000).

Water age in the reservoir can be difficult to determine if the reservoir is not studied. The older the water, the more stagnant the water is, for a more extended period, allowing for microbial growth (Rossman and Grayman, 1999). Increased water age can be due to the design of the reservoir, for example, placement of the inlet and outlet and the height to diameter ratio of the inlet and outlet (Rossman and Grayman, 1999). Parameters related to water age can be traced back to the chlorine disinfection residual; the less chlorine disinfection, the older the water age.

The inlet size and configuration within the reservoir can determine the quality of the water and influence microbial growth. Water entering the reservoir can be considered to flow in a jet-like circular motion (McNaughton and Sinclair, 1966). The current created by the inlet needs to be adequate to allow mixing to occur. Configuration of the different inlets can be problematic. For example, a tangential inlet will create dead spots in the middle of the reservoir; inlets directed at the wall will not be able to develop a jet-like motion. They will therefore not allow for mixing, and a large diameter inlet will create low velocities in the water and again, there will be no mixing (Grayman *et al.*, 2004). All the mentioned examples will allow for microbial regrowth.

Understanding all the parameters and properties playing a role in the reservoir would conceive a computational fluid dynamics model (CFD). The model is based on the physical process governing the fluid flow in the reservoir. The model will show the design and operation of the reservoir while allowing for predictions of how a change in operations and physical characters affect the reservoir (Grayman *et al.*, 2000). The CFD model allows for a visual representation of the characteristics accurately to better understand the system. Understanding the CFD would allow reasoning for the microbes present in the reservoir and possibly explain their functionality. The model can be created by evaluating the design, such as inlets, positioning, and sampling temperatures from different depths.

A study done by a colleague, S. MacRae, showed increased abundance levels in microbial communities in reservoirs in a drinking water system. Identification of the microorganisms found in reservoirs and their functions could aid in the management of reservoirs in the drinking water system. All the interactions between microbes and their environment and conditions in the reservoir could be explained by understanding and investigating possible functional networks, which would explain why microbes survive and can increase.

In this portion of the study, permission was obtained a local municipality in Gauteng to include one of their community reservoir in this study. Reservoir A was situated in an area known to experience water quality problems. Water leaving the purification plant is typically of acceptable quality but by the time it reaches the consumer the water quality may have deteriorated significantly. This reservoir was sampled over several months to determine the interplay between design and flow patterns of the reservoir on the microbial quality within the reservoir.

With the help of the Department of Civil Engineering (UP), measurements such as flow, temperature and water levels within the reservoir were collected as input to create a crude model of the hydrology within the

reservoir using the Computer Fluid Dynamics (CFD) system. This information was used to identify possible stagnation zones in the reservoir. This data served as input for the location of sampling points within the reservoir. To ensure that enough biological material was collected, all the samples were concentrated with membrane filtration. This was followed by DNA extraction from each sample which was used for 16S profiling and metagenome studies.

3.2 MATERIALS AND METHODS

3.2.1 Reservoir selection and design

The first part of the study entails the investigation of a community reservoir (Reservoir A) in Gauteng. The reason for the chosen reservoir is that the water quality is questionable. Water leaving water purification plants are typically of acceptable quality but when it reaches the consumer the water quality is significantly lower (Liu *et al.* , 2013). It is suggested that service reservoirs do not provide adequate mixing and that the residence time influences the water quality. The scope of the study is to determine the problem areas inside the reservoir and propose possible changes to the design that will improve the mixing which ultimately contribute to the quality of the water reaching the community. Water from the treatment plant would typically be supplied to the community as shown in Figure 3.1. However, the pressure tower is not in use and the water is directly transported from the service reservoir to the community.

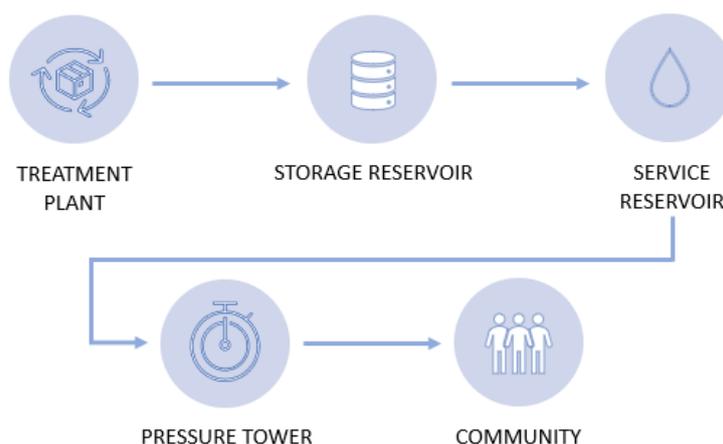


Figure 3.1: Water transport from treatment plant to the community

Reservoir A that was used for the study is a circular ground storage reservoir with a single inlet and outlet. No as-built drawings were available, and a visual inspection was thus performed by measuring the inlet, outlet, and overall characteristics of the service reservoir. Characteristics of the reservoir is summarised in Table 3.1. The inlet is elevated above the ground level directed horizontally to the reservoir.

Table 3.1: Reservoir characteristics for the study reservoir

Reservoir characteristic	Dimensions (m)
Reservoir Height	8
Reservoir Diameter	30
Inlet diameter	0.3
Outlet diameter	0.3
Reservoir Columns	0.4 x 0.4
Air vent diameter	0.3
Manhole size	1 x 0.7

The outlet is situated at the bottom centre of the reservoir at ground level. There are 32 square support columns in the reservoir with two manhole covers on either side of the reservoir. Six air vents are also present on the roof of the reservoir. The reservoir with the inlet and outlet location is depicted in Figure 3.2 with the inlet characteristics are shown in Figure 3.3.

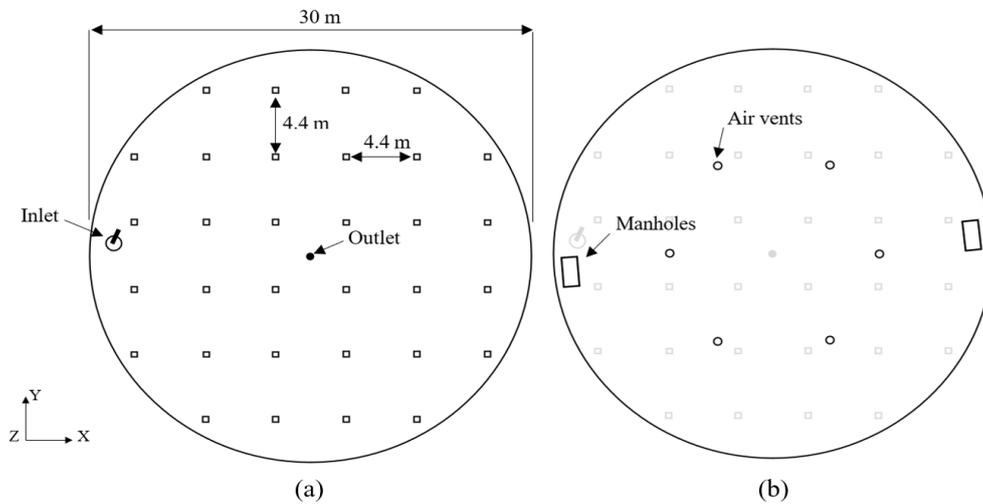


Figure 3.2: Reservoir outlay as seen (a) inside and (b) from the top.

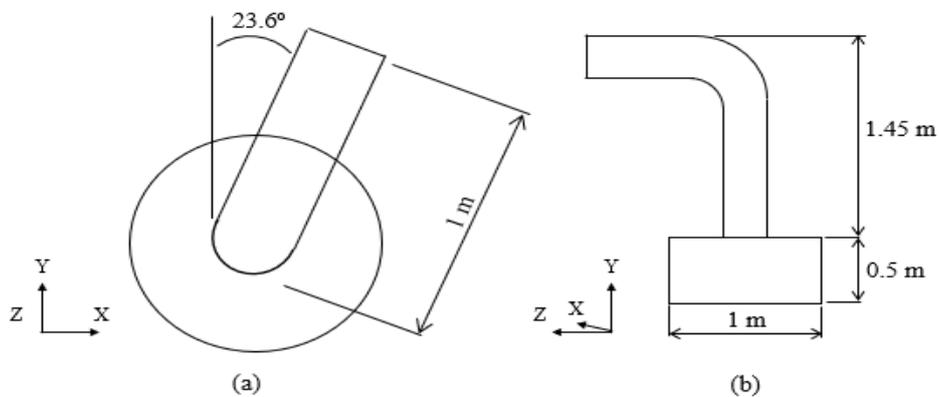


Figure 3.3: Detailed diagram of (a) inlet and (b) front view of inlet.

3.2.2 Analysis of Reservoir A flow dynamic

Analysing mixing in a reservoir can be described with the use of the methods shown in Figure 3.4. Water velocities can be obtained through costly methods to determine where water forms stagnant zones. CFD has been widely used in place of the expensive measuring techniques to obtain flow fields. Computational fluid dynamics (CFD) is a cost-effective tool that can be used to determine the flow patterns in reservoirs (Martínez-Solano *et al.*, 2010). Khan *et al.* (2012) assessed the CFD capabilities of simulating flow patterns and velocity distributions in storage tanks and found that the CFD results are an accurate representation of the experimental results. The advantage of a CFD analysis is that it allows for flow patterns to be simulated in a reservoir before it is constructed. It also enables the provision of detailed information such as the flow structure, velocity distributions and components at any point in the reservoir (Khan *et al.*, 2012).

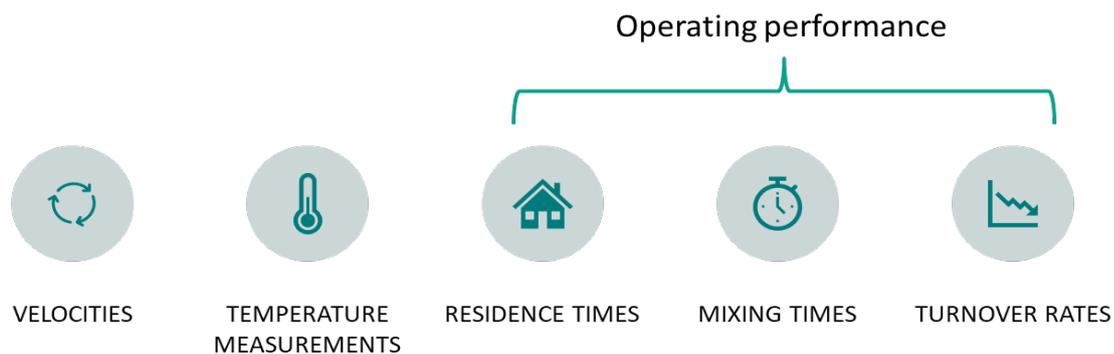


Figure 3.4: Methods for describing mixing efficiency in service reservoirs.

Temperature measurements are an effective technique to determine if water stratifies in a system as water coming into the reservoir can be colder than the bulk water already in the reservoir. The operating conditions of the filling and drawing cycles of a reservoir can also influence the mixing efficiency. The operating conditions can be analysed through the residence times, mixing times and turnover rates. The time it takes for a tracer that was injected to a body of water to become uniform can also be used to determine the mixing characteristics. This is called the mixing time. Turnover rates (water entering and leaving the system) must also be closely evaluated to ensure efficient turnover.

The following steps were followed so far to determine whether good mixing is present in the reservoir:

- A thorough literature study was conducted to study the influence of mixing on the water quality as well as the effects of tank design on mixing.
- The reservoir characteristics were studied to determine the tank diameter, column size and position, and reservoir height. A scanned representation of the reservoir was obtained.
- Temperature measurements were recorded from December 2019 to December 2020 using thermistors. Thermistors were placed at 48 different locations and heights to determine the possibility of stagnation.
- A pressure probe was used to measure the water level from December 2019 to December 2020. This data was used to obtain theoretical residence times due to the operating conditions of the reservoir.
- Flow rates at the inlet pipe near the reservoir was obtained using an ultrasonic flowmeter for use in the CFD model.
- A CFD CAD model was set up to model the flow patterns in the reservoir.
- CFD models for three different optimisation results were set up and will be run once the base case is verified.
- CFD validation is currently taking place.

3.2.2.1 *Determining the mixing characteristics of the reservoir*

The full-scale reservoir was used to gather field data to determine the mixing characteristics of the reservoir. Temperature variations in the reservoir were measured with temperature sensors that were placed at various locations in the reservoir. The inflow rate, outflow rate and water levels in the reservoir were also measured.

3.2.2.2 *Temperature measurements*

Forty-eight temperature measurements at 16 locations were used to determine the variation in temperature inside the reservoir. The equipment included temperature sensors which were attached to cables at 3 depths

inside the reservoir. The cables, suspended from ropes which spanned between the columns, were weighted down with a concrete block to ensure that the cables remained intact during the measurement period. The temperature sensors were connected to a data logger which stored the temperature data. The setup of the temperature sensors at various depths are shown in Figure 3.5. The locations of the sensors in the reservoir with their corresponding numbers are depicted in Figure 3.6. The CR300 data logger with its setup is shown in Figure 3.7.

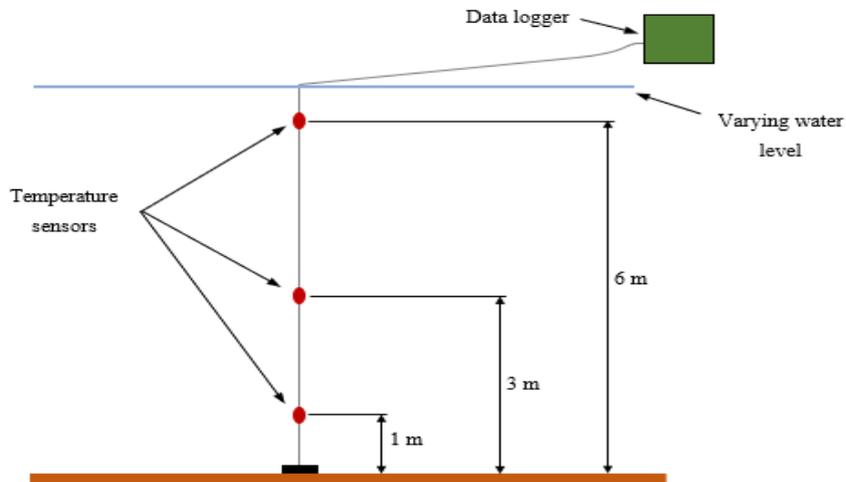


Figure 3.5: Depths of the thermistor as installed in the reservoir.

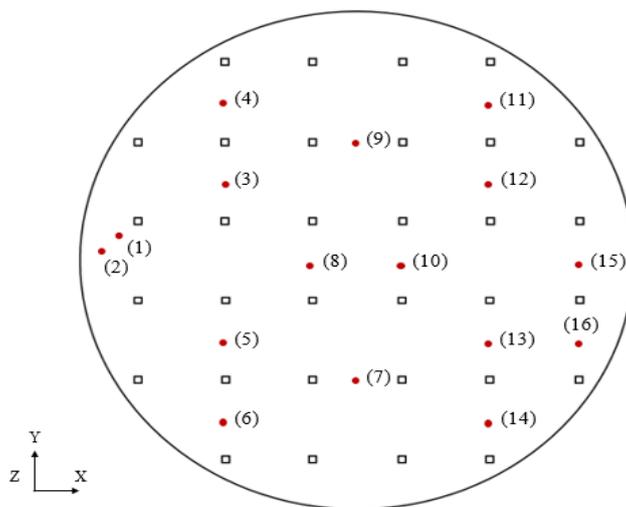


Figure 3.6: Thermistor locations within the reservoir.

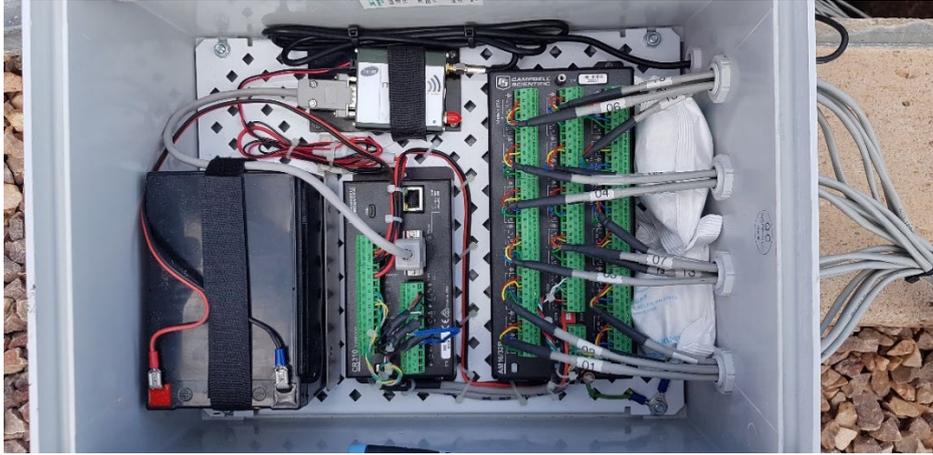


Figure 3.7: Photo of the CR300 data logger used during the study

The accuracy of a thermistor is dependent on the thermistor interchangeability, the bridge-resistor accuracy and the Steinhart-Hart equation (Campbell Scientific, 2018). A 1-point or 2-point calibration can be carried out to find the offset parameters for each thermistor. The thermistors were calibrated by placing the thermistors at the same location in the reservoir. A reference thermistor was used, and it was assumed that the thermistor measured an accurate temperature reading. Each thermistor's offset with the reference thermistor was then calculated and used for calibration. The calibration was completed to acquire relative temperatures and not absolute temperatures. The temperature sensor specifications are summarised in Table 3.2.

Table 3.2: Temperature sensor information (Campbell Scientific, 2018)

Specification	°C	
Measurement range	-35 to 50	
Accuracy	Worst case	± 0.4 (-24 to 48)
		± 0.9 (-35 to 50)
	Interchangeability error	± 0.1 (0 to 50)
Steinhart-Hart equation error	± 0.01 (-35 to 50)	

3.2.2.3 Water Levels and Inflow Rates

The inflow rate was determined using an ultrasonic flowmeter. The ultrasonic flowmeter is a convenient method for determining the flow in a pipe due to its non-intrusive measurement technique. The distance between the sensors of the flowmeter was determined and the flowmeter was placed at a straight section on the inlet pipe. The ultrasonic flowmeter on the inlet pipe is shown in Figure 3.8.



Figure 3.8: Photo of the ultrasonic flowmeter used during this study.

The study required the water elevation in the reservoir to determine residence times and whether the temperature sensors were submerged. The water level was continuously monitored using a submersible pressure probe. Pressure probes that store information internally were used and data from the sensors were downloaded onto a computer after the study. The water level was used to determine information such as:

- Time at the start of a filling cycle
- Water level at the start of a filling cycle
- Time at the end of a filling cycle
- Water level at the end of a filling cycle

3.2.2.4 Computational setup

The commercial CFD software package STAR CCM+ was used to model the flow in the reservoir. This section describes the approach followed to set up the model. The geometry for the model was set up according to the dimensions and specifications described above. Optimisation design methods that were considered to improve the mixing characteristics in the reservoir are also defined in this section.

3.2.2.5 Geometry

The geometry for the model was set up according to the dimensions and specifications in shown above. The CFD CAD model is presented in Figure 3.9 and Figure 3.10.

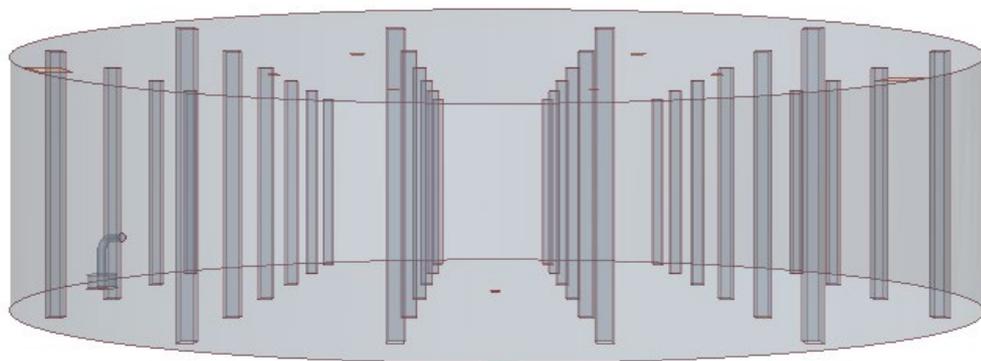


Figure 3.9: CAD model of the side view of the reservoir.

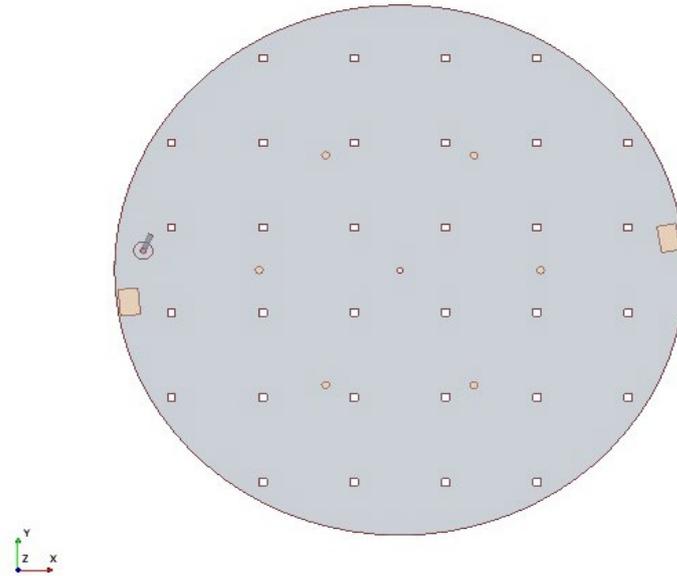


Figure 3.10: CAD model of the top view of reservoir.

3.2.2.6 Boundary and initial conditions

Boundary conditions to guide flow paths and define regions were determined and is summarised in Table 3.3.

Table 3.3: Boundary conditions used to guide flow paths and define regions

Boundary condition	Description	Model
Velocity inlet	Known inlet flow	Inlet
Pressure outlet	Allow air to escape as water fills the reservoir	Air vents
		Manholes
Negative Mass flow inlet	Known outlet flow used as negative inlet	Outlet
Walls	Specify wall as solid with no fluid escaping	Reservoir walls and internal columns
		Bottom of reservoir
		Reservoir roof
		Inlet and outlet walls

3.2.2.7 Near-wall Modelling

Considering high Reynolds numbers in reservoirs and the focus to be more on the mixing processes in the reservoir and not the forces on the wall, wall functions were used to account for the near wall modelling in the study.

3.2.2.8 Physics models

The physics models chosen for the simulation is depicted in Table 3.4. A discussion of the choice of physics models will be presented further in this chapter. For this model, unsteady flow is used and the water flow in the reservoir is turbulent.

Table 3.4: Physics models used for simulation

Physics model	Description	
Domain	Three dimensional	
Gravity	Incorporated	
Time	Implicit unsteady	
Material	Eulerian multiphase	Air
		Water
Turbulence model	Realizable k-ε model	
Free surface	Volume of fluid	HRIC scheme
Segregated flow	Segregated multiphase temperature	

3.2.2.9 Turbulent flows

There are two two-equation turbulence models in the RANS equations set namely the K-Epsilon (k-ε) and K-Omega (k-ω) models. In the case of determining the flow characteristics in service reservoirs, the k-ε model is more appropriate and was used in the evaluation of tank mixing in previous studies (Moncho-Esteve *et al.*, 2015; Xavier and Janzen, 2017; Montoya-Pachongo *et al.*, 2016; Yeung, 2001; Swayne *et al.*, 2010b; Moncho-Esteve *et al.*, 2013).

3.2.2.10 Multiphase flow

The Volume of Fluid (VOF) model was selected together with the Eulerian method. This approach is appropriate for the study as there are distinct respective regions of air and water. The HRIC (High Resolution Interface Capturing) was utilized as it is used to trace the interface between the air and water section over time.

3.2.2.11 Segregated multiphase temperature

Temperature studies were used to validate the CFD model. It should be noted that using only a numerical approach to determine the flow patterns cannot be deemed as accurate. Validating a CFD model can sometimes be more consuming than the model itself. This is due to the complex nature of turbulent flow, three-dimensional flow and low velocities in service reservoirs (van der Walt, 2002). In the model, segregated flow with multiphase temperature was used to analyse the temperature variations in the reservoir. This included buoyancy. Buoyancy was included in the study by adding gravity and a thermal expansion coefficient as field functions. Thereafter buoyancy was added with the field function using Equation 3.1.

$$\rho * \beta * (T_w - T_s) * g \quad \text{(Equation 3.1)}$$

Where:

T_w = Initial water temperature (°C)

T_s = System temperature (°C)

An initial base case was considered for validation. This assumed a uniform air temperature in the reservoir, a uniform temperature inside the reservoir and, a constant temperature for the inflowing fluid. This resulted in the initial conditions as shown in Figure 3.11. The reservoir walls were also considered as part of the temperature verification. To model the temperature influence of walls, the heat flux can be calculated and used in the model.

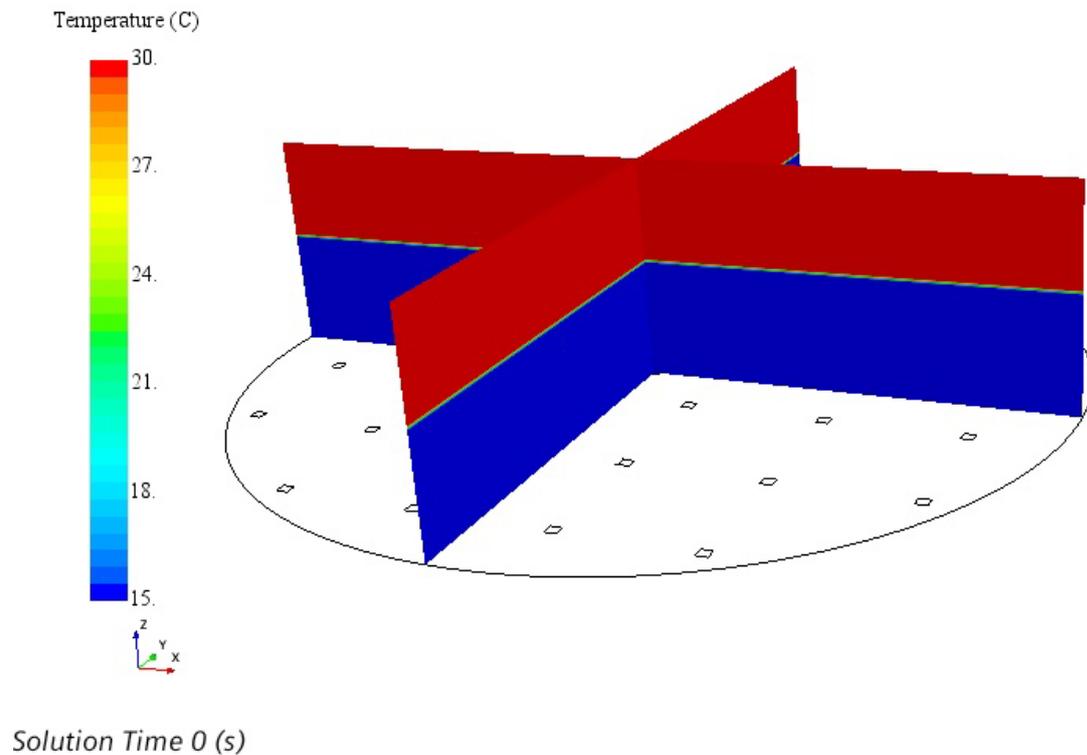
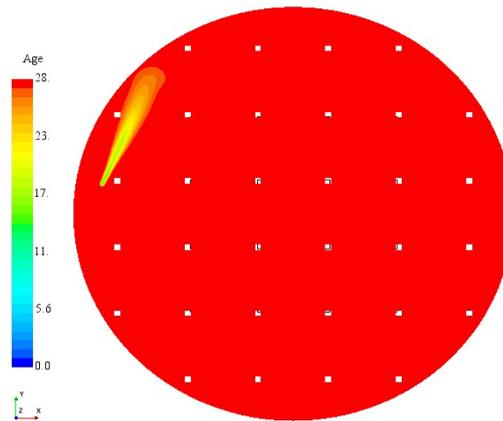


Figure 3.11: Initial temperature conditions

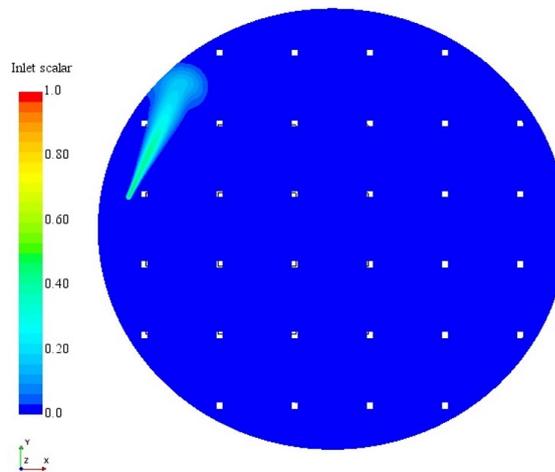
3.2.2.12 Tracer Methods

A passive scalar illustrating the water flowing into the reservoir was used to track the inflow during the time spent in the system. The initial conditions of the reservoir were set such that no tracer was initially in the system. The passive scalar was added to the inlet boundary such that 100 % of the inflowing water represents the tracer. The inflowing water was then tracked to determine the mixing inside the reservoir. Another passive scalar was added to track the mean age of the water in the reservoir. The properties of the scalar were modified to ensure that advection dominates. A turbulent Schmidt number of 1 was chosen for the study as the scope of the study did not include the accurate prediction of these numbers. The passive scalars described are presented in Figure 3.12 and Figure 3.13.



Solution Time 28.2 (s)

Figure 3.12: Age passive scalar based on the model.



Solution Time 28.2 (s)

Figure 3.13: Inlet passive scalar based on the model.

3.2.3 Sample collection for microbial profiling and functional analyses

The sites within Reservoir A from which samples were collected for the microbial community analyses are shown in Figure 3.14.

Sampling was conducted over several months, which included March, August, September, October, November, and December 2020. Samples were collected at designated sample points seen in Figure 3.14, within the reservoir and at different depths in replicate. Samples points included five points within the reservoir at different depths (i.e. 1 m and 3 m from the ground), as well as the inflow and outflow. The treated water is extracted from a large dam, in which the water undergoes conventional treatment, i.e. coagulation, flocculation, sedimentation, carbonation, filtration (i.e. RS filtration) and disinfection (chlorination). Approximately 8L of each sample was collected in sterile 8L Large Narrow Mouth Nalgene polycarbonate bottles (Thermo Scientific™, South Africa) and transported on ice to the laboratory, and stored in a large refrigerator. The samples were concentrated using STERIVEX™-GP 0.22 µm polycarbonate membrane filter unit (Merck Millipore, South Africa). Filtration was facilitated using a Watson-Marlow 101U peristaltic pump and sterile Nalgene™ 50 Platinum-Cured Silicone Tubing (Thermo Scientific™, South Africa). After filtration the polycarbonate filter membranes were transferred and stored in

sterile microcentrifuge tubes at -20°C until DNA extraction (Kwon *et al.*, 2011; Pinto *et al.*, 2012).

The same samples collected from Reservoir A for the 16S rRNA community profiling were used in this study. Sampling was conducted over several months as described above. It included March, August, September, October, November, and December 2020. Samples points included five points within the reservoir at different depths (i.e. 1 m and 3 m from the ground), as well as the inflow and outflow. The collection and processing of the samples as well as the extraction of DNA are described in detail in Section 3.2.4.

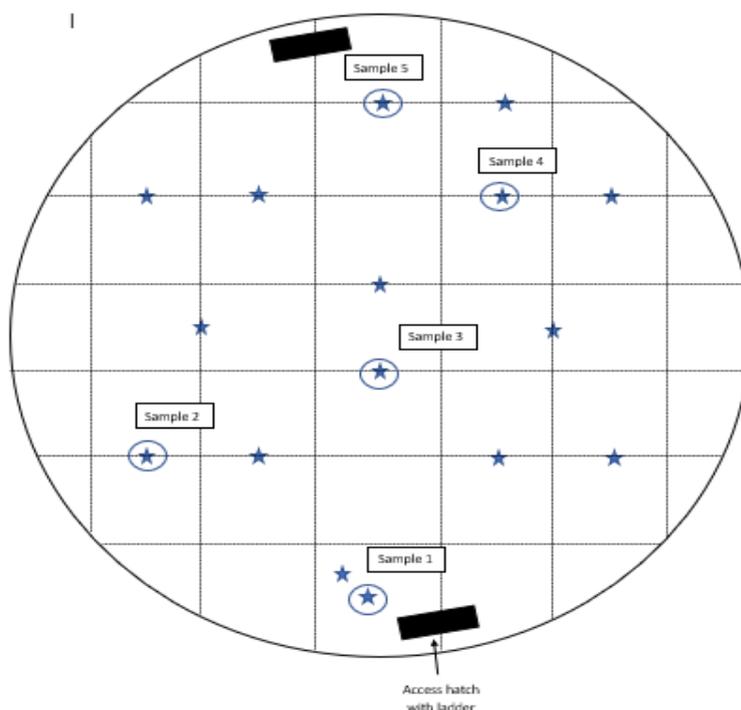


Figure 3.14: Schematic diagram representing the sampling points at 1 meter and 3 meters in Reservoir A.

3.2.4 Methods for microbial profiling of samples from Reservoir A

3.2.4.1 Separation of DNA fragments

The Sterivex filter membranes were cut into several pieces using a scalpel sterilized with ethanol. The cut-up membranes were then placed into a 2 ml Lysing Matrix Tube E using ethanol sterilized tweezers. A volume of 300ul of 2X TENS (100mM Tris-HCl [pH 8.0], 40 mM EDTA, 200 mM NaCl, 2 % SDS) buffer was added and vortexed vigorously. A volume of 1 ml of phenol:chloroform:isoamyl alcohol (25:24:1, pH 8) was added and bead beaten at maximum speed for 40 seconds. The tubes were then centrifuged at 14 000 x g for 10 minutes. A volume of 200 ul of 2 X TENS buffer was added and the tubes were bead beaten and centrifuged again as mentioned above. This was repeated one more time. A volume of 700 ul of phenol:chloroform:isoamyl alcohol (25:24:1) was added and mixed thoroughly by repeated inversions (this was not vortexed). Once mixed, it was then centrifuged at 14 000 x g for 5 minutes. The aqueous phase was then transferred to phase lock gel tube (previously, the 2.0 ml Heavy Phase Lock Gel tubes was centrifuged at 14 000xg for 1 minute. 2 tubes per sample).

A volume of 350 μ l of 7.5 M NH_4Ac was added and mixed with repeated inversions. A volume of 600 μ l of chloroform was added and mixed thoroughly by repeated inversions. The tubes were then centrifuged at 14 000 x g for 5 minutes. The aqueous phase was then transferred to 2.2 ml microcentrifuge tubes. The nucleic acid was precipitated from the final collected aqueous phase by adding 600 μ l of isopropanol and 6 μ l of 15 mg/ml Glycoblue, mixed gently, and incubated at 80°C for 10 minutes. The tubes were taken out and centrifuged for 30 minutes at 12 000 x g. The supernatant was carefully removed while keeping the blue pellet in the tube. The precipitate was washed with 1 ml of 80 % ethanol and then centrifuged again for 30 minutes at 12,000 x g. The supernatant was carefully removed again, and the tubes left to air dry for 5-10 minutes. The samples were resuspended in 50 μ l of DNase free water and then pipette a volume of 25 μ l into a second vial. A gel was run to ensure successful amplification Gel electrophoresis involves 0.5g of agarose gel being added to 1 x TAE buffer (100ml of 50 x TAE to 4.9L of water = 5L of 1 x TAE) to produce a 1% gel. This was heated in a microwave for a few minutes until the agarose is completely dissolved. Once cooled down it was poured into the electrophoresis tray, the comb was placed in agarose make the lanes and left to set. The lanes were loaded (1 μ l of loading dye was added to 2 μ l of DNA) and the gel was left to undergo electrophoresis for 30 minutes at 80 V.

3.2.4.2 16S rRNA amplification

A 16S rRNA PCR was performed to ensure that the DNA is of good quality and that there is successful amplification before samples are sent to the University of Michigan, USA for Illumina Miseq sequencing. The 16S rRNA gene was amplified using two universal primers: 27F (5'-AGA GTT TGA TCC TGG CTC AG-3') and 1492R (5'-GGT TAC CTT GTT ACG ACT-3') adapted from Edward *et al* (1989). The polymerase chain reaction (PCR) reactions were carried out using the BIO-RAD T100™ Thermal Cycler. The PCR mixtures that contained 10 x reaction buffer, 1.5 mM MgCl_2 , 250 μ M of each nucleotide (dATP, dCTP, dGTP, dTTP), 10 pmol of each primer (forward and reverse), 1.5 μ l Taq DNA polymerase, 16.85 μ l nuclease free water (Qiagen, South Africa) and 0.5 μ l of the extracted genomic DNA giving a final volume of 25 μ l. The PCR cycling conditions consisted of an initial denaturation step at 92°C for 10 minutes, followed by 30 cycles of denaturation at 92°C for 1 minute, annealing at 58°C for 1 minute, extension at 75°C for 1 minute, and a final extension at 75°C for 5 minutes. Depending on the obtained DNA concentrations following the phenol-chloroform DNA extraction method, an additional PCR was done with DNA dilutions (10-fold or 100-fold) to improve amplification and prevent 16SrRNA amplification bias. The resulting amplicons were then subjected to agarose gel electrophoresis (1% agarose gel), where the samples were run for 30 minutes at 80 V, to observe successful 16S rRNA gene amplification.

3.2.4.3 16S Sequences processing

Data analysis was done using software package DADA2 version 1.12. The DADA2 package is centred around the DADA2 algorithm for accurate high-resolution of sample composition from amplicon sequencing data. The DADA2 algorithm is both more sensitive and more specific than commonly used OTU methods and resolves amplicon sequence variants (ASVs) that differ by as little as one nucleotide. Features of the program include quality filtering, dereplication, learn error rates, sample inference, chimera removal, merging paired reads and taxonomic classification.

3.2.4.4 Illumina Miseq sequencing

Samples were pooled according to their sampling months and sent for Illumina Hiseq sequencing (2 x 150bp and \pm 80 million reads per sample) to the Agricultural research council (ARC) Biotechnology platform, South Africa. The sequences were then produced using a metagenomics pipeline, SqueezeMeta (Tamames & Puente-Sanchez, 2019), which creates bins using a co-assembly approach. SqueezeMeta is a fully automated pipeline that allows for assembling related metagenomes and retrieving the individual genome via

binning procedures. The program uses; Trimmomatic (Bolger *et al.*, 2014) for filtering, Megahit (Ayling *et al.*, 2019) for assembly, Bowtie (Langmead *et al.*, 2009) for mapping, COG/KEEG/Pfam (Bose *et al.*, 2015) for annotation, MaxBin/Metabat (Wu *et al.*, 2015; Kang *et al.*, 2019) for binning and GTDB_TK (Chaumeil *et al.*, 2019) online database for taxonomic classification. Retained bins were classified using SqueezeMeta, MiGA and GTDB-Tk databases. Annotated and classified bins were then analysed in KEGG Decoder (Graham *et al.*, 2018) and KEGG Mapper (Kanehisa & Goto, 2000).

Another batch of DNA samples were sent to the University of Michigan, USA for Illumina Miseq sequencing. The Illumina sequencing platform was used to identify and characterize the microbial community. Amplification primers targeting the V4 hypervariable region of the 16S rRNA gene were used as described by Kozich *et al* (2013). The gene was amplified from each sample using the Dual-indexing sequencing strategy developed by Dr. Patrick D. Schloss. The sequencing reads were 250bp long and have complete overlap.

3.2.5 Community comparisons

R studio version 3.2.2 software was used to determine the alpha and beta diversity of the bacterial communities within each sample. The program conducted the required statistical analyses and compiled different figures and heatmaps to display the information. Alpha diversity was measured using three different parameters: Species observed (Sobs), Shannon diversity index (H') and Pielou (J). Sobs is the number of species that is observed within each sample point and provides a good indication of richness in the system. H' examines diversity within the sample point by examining presence, absence and abundance. J examines evenness within the sample points by examining numbers and abundance of species. Beta diversity will examine the differences in bacterial composition among all sites by creating a distance matrix (a square matrix containing distances, this is a pairwise analysis between the two groups of interests, showing similarity or differences). Bray-Curtis determines community structure (incidence and abundance) while Jaccard determines community membership (absence and presence). Values closer to 1 indicate high dissimilarity among the community (Jaccard and Turrisi, 2003; Bray and Curtis, 1957).

3.2.6 Flow cytometry analysis

Flow cytometry was performed to determine cell counts and to quantify the intact and damaged bacterial cells in each sample (Berney *et al.*, 2007). The water samples will be stained with SYBR Green I and Propidium Iodide (SGPI), as it allows for the differentiation between intact and damaged cells by means of a fluorescent light, using a flow cytometer (Nescerecka *et al.*, 2015). For this analysis, 500 µl of the water sample will be put into a tube (this will be done in duplicate for each water sample). To the one tube 5 µl of SG/PI will be added (6 µl PI final concentration) and vortexed and to the second tube 5 µl of SYBR Green I (SG) will be added instead. The samples will then be incubated in the dark at 37° C for 10 minutes. The samples will be vortexed and viewed used a flow cytometer.

3.3 RESULTS AND DISCUSSION

3.3.1 Assessment of current design and flow patterns of the Reservoir A

The current mixing patterns in the reservoir was assessed using temperature measurement. The filling and drawing cycles were also noted to determine the residence time of the water inside the reservoir. The mixing patterns and residence times were assessed to determine whether the reservoir is designed optimally.

3.3.1.1 Operating conditions

The water levels and the inflow and outflow rates were assessed to determine if the reservoir is operating optimally. This includes the turnover rate and the mixing time of the water in the reservoir.

3.3.1.2 Fill-draw cycles

The water levels were analysed to determine the fill and draw cycles of each day. An example of the typical fill-draw cycles for the reservoir is illustrated in Figure 3.15. The start of the filling cycle is the end of the previous drawing cycle while the end of the filling cycle is also the start of the next drawing cycle. As seen in Figure 3.15, the fill-draw cycles ranged from one to two cycles per day. The days with two fill-draw cycles were divided into an early fill-draw cycle and a late fill-draw cycle. The early fill-draw cycle occurred in the mornings and the late fill-draw cycles in the afternoon each day. For days with only one fill-draw cycle, the start of the filling cycle followed in the morning and the end of the drawing stage occurred the following day in the morning, with some exceptions.

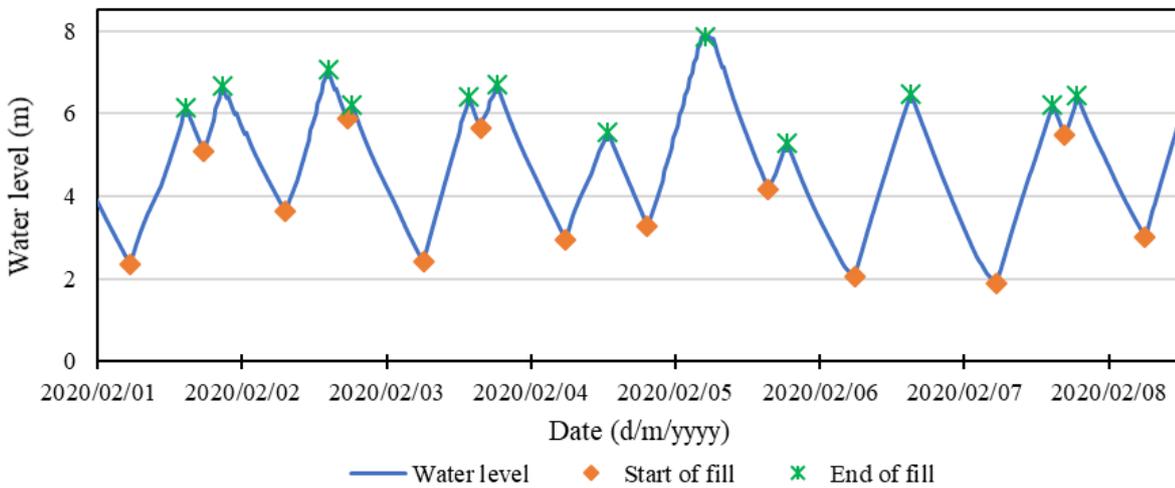


Figure 3.15: Water levels linked to the fill-draw cycles.

3.3.1.3 Flow rates

The inlet velocities obtained using the ultrasonic flowmeter are shown in Figure 3.16. Outflow rates could however not be obtained using the flowmeter. Water levels and the inflow rates were used to obtain the outflow velocity. These flow rates were used in the CFD simulation.

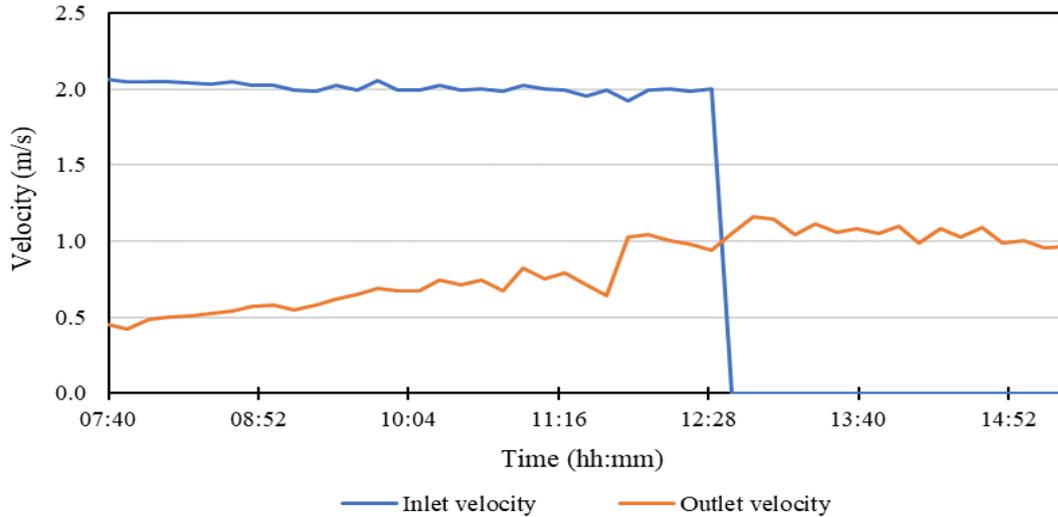


Figure 3.16: Inlet and outlet velocities.

3.3.1.4 Turnover rates

In order to minimise the water age in a reservoir, the amount of turnover must be sufficient (Martel *et al.*, 2002). Turnover refers to water volumes that are exchanged from filling and drawing cycles. The turnover rate can be described in two ways as stated by Kirmeyer *et al.* (1999):

- Average time, also called the detention time that the entire volume of water spends in the reservoir; or
- The percentage of the tank volume that is replaced per day.

The turnover rates were described using the detention time and theoretical residence time from Equation 3.2 and 3, respectively.

$$\text{Detention time} = \left(0.5 + \frac{V}{\Delta V}\right)(t_d + t_f) \quad \text{(Equation 3.2)}$$

Where:

V = Volume of water at the start of the filling period

ΔV = Change in water volume during filling period

t_d = Draw time

t_f = Fill time

$$\text{Theoretical residence time} = \frac{V_{\max}}{V_{\max} - V_{\min}} \frac{1}{N} \quad \text{(Equation 3.3)}$$

Where:

V_{\max} = Maximum daily volume

V_{\min} = Minimum daily volume

N = Number of fill-draw cycles

The theoretical residence times from 1 January to 6 December 2020 are depicted in Figure 3.17. Theoretical residence times for two fill-draw cycles and one fill-draw cycle were analysed. Days where two fill-draw cycles were observed resulted in lower residence times than days where only one fill-draw cycle was observed. The theoretical residence times for some days with one fill-draw cycle were more than double that of the days with two fill-draw cycles. The average residence time for days with two fill-draw cycles were 20.8 hours and 40 hours for days with one fill-draw cycle.

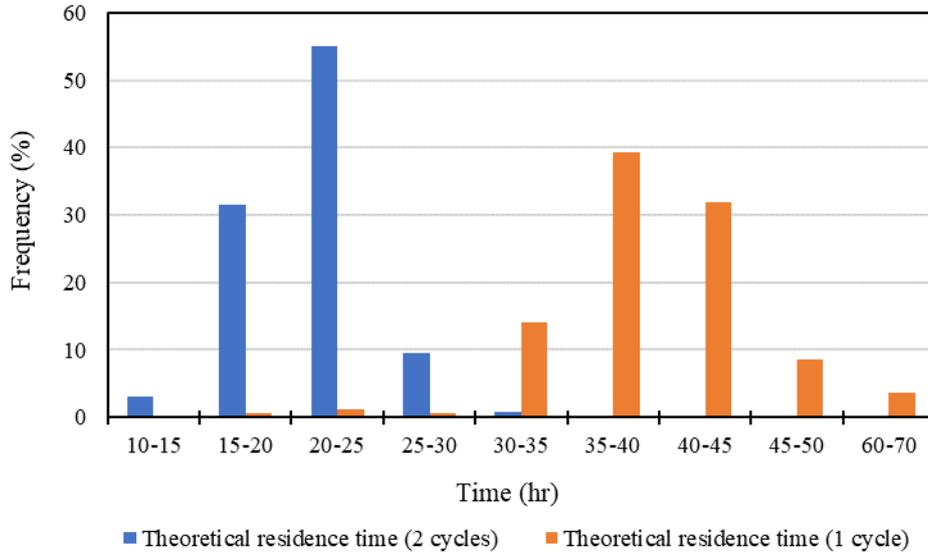


Figure 3.17: Theoretical residence time for 1 January to 6 December 2020.

For days with one fill-draw cycle, the detention time ranged from 20 hours to 70 hours with an average of 27.9 hours as shown in Figure 3.18. The detention times for days with one fill-draw cycle was larger than the early fill-draw cycle but generally less than the late fill-draw cycle. For days with two fill-draw cycles, the early fill-draw cycle commonly resulted in lower detention times than the late fill-draw cycle. The early fill-draw cycle had an average detention time of 15 hours with a maximum of 42 hours. The late fill-draw cycle had a maximum detention time of 275.31 hours which are one of two outliers in the study period. The average including the outliers were 78.75 hours. The days where the detention time differed significantly was on 29 January and 2 February. On these days, the second filling time was under an hour. If these cases were handled as if there was only one fill-draw cycle, these values would change as shown in Table 3.5. This shows that the choice of whether a small filling cycle can be considered as a separate full filling cycle must be re-evaluated.

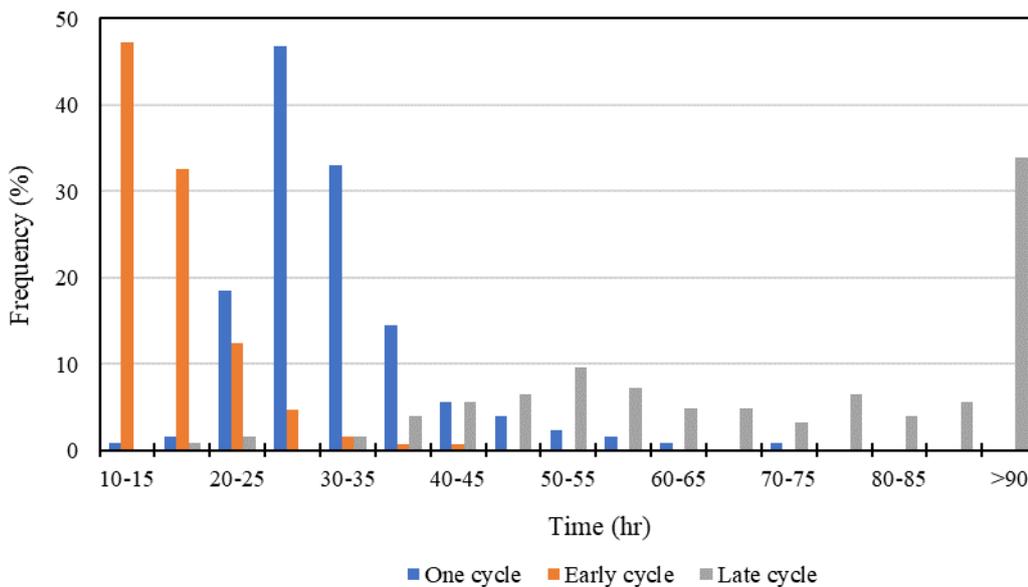


Figure 3.18: Detention time for the period of 1 January to 6 December 2020.

Table 3.5: Change in fill-draw cycles during the study

	Filling cycle	29 January	2 February
Detention time (hr)	Two fill-draw cycles	275.31	272.66
	One fill-draw cycle	32.55	44.23

The detention times for days with two fill-draw cycles, the late and early fill-draw cycles varied considerably. A day with two fill-draw cycles is summarised in Table 3.6. For the early fill-draw cycle, the filling period was longer than the drawing period whereas the opposite is observed for the late fill-draw cycle. Fill-draw cycles with longer drawing periods tended to have much higher detention times due to the large volume exchange during this period, which is not factored into Equation 3.2.

Table 3.6: Fill-draw information of a day with two fill-draw cycles

	Early fill-draw cycle		Late fill-draw cycle	
	Period (hr)	Volume exchange (m ³)	Period (hr)	Volume exchange (m ³)
Filling period	8.5	2 100.111	3.67	537.818
Drawing period	3.5	842.754	8	1 649.769
Detention time (hrs)	17.716		78	

3.3.1.5 Mixing times

Mixing times can be described as the time it takes for a tracer to become uniform and refers to the recommended filling time during a fill-draw cycle. The formula for the mixing time (t_m) is shown in Equation 3.5 and can be used with any consistent set of units (Rossman and Grayman, 1999; Roberts *et al.*, 2006).

$$\tau_m = t_m \frac{M^{1/2}}{V^{2/3}} \quad \text{(Equation 3.5)}$$

With:

$$Q = \frac{\pi}{4} d^2 u \quad \text{(Equation 3.6)}$$

$$M = uQ \quad \text{(Equation 3.7)}$$

Where:

Q = Inflow rate

u = Velocity

d = Inlet diameter

t_m = Mixing time

τ_m = Dimensionless mixing time

V = Volume of the water

The recommended τ_m values for different H:D ratios are shown in Equation 3.8 (Tian and Roberts, 2008).

$$\tau_m = \begin{cases} 10, & \frac{H}{D} \leq 0.8 \\ 10.0 + 3.5 \left(\frac{H}{D} - 1 \right), & \frac{H}{D} > 0.8 \end{cases} \quad \text{(Equation 3.8)}$$

The mixing times and filling times for one fill-draw cycle are shown in Figure 3.19, and two fill-draw cycles in Figure 3.20. This shows that the filling time well exceeds the recommended mixing time to achieve 90 % mixing in the tank with some days in the two fill-draw cycles where the recommended mixing time was not achieved.

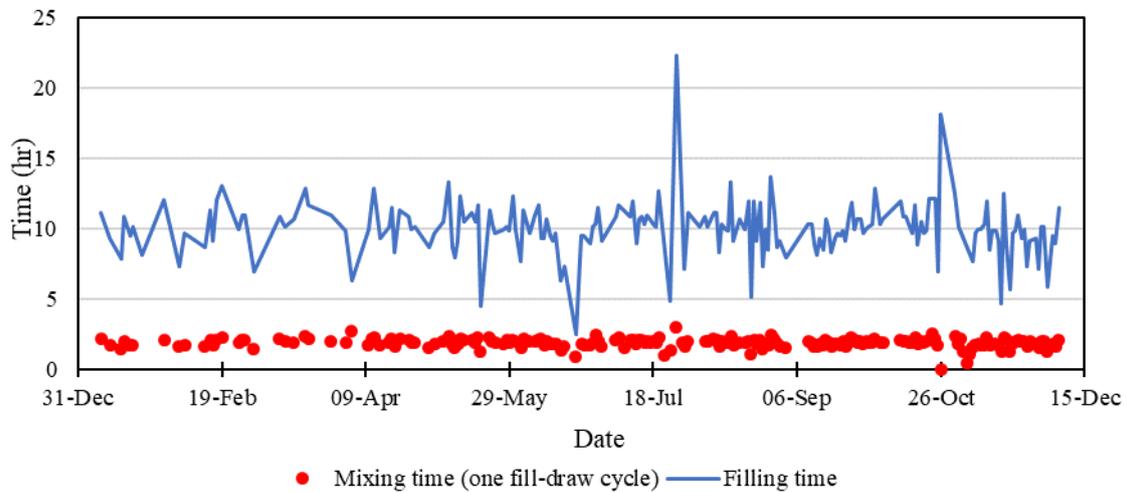


Figure 3.19: Mixing times as calculated for one fill-draw cycle.

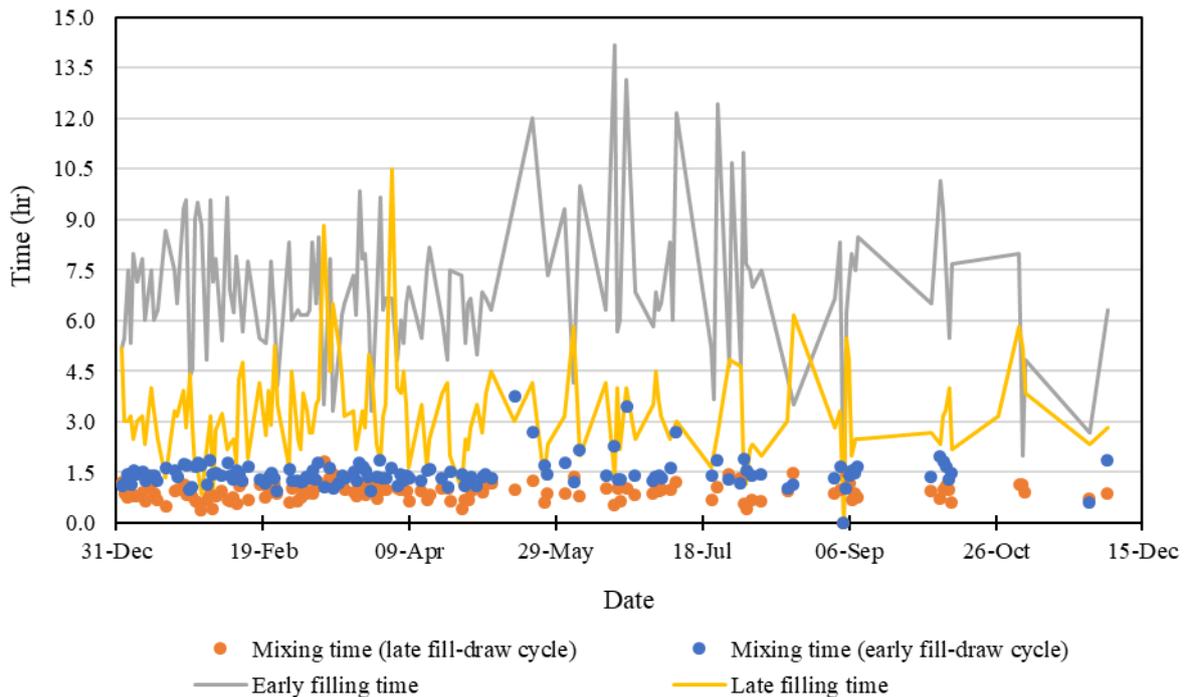


Figure 3.20: Mixing times as calculated for the early and late fill-draw cycle.

3.3.1.6 Volumetric exchange

The volume exchange to achieve 90 % mixing in a tank can be calculated using Equation 3.9 and the results are shown in Figure 3.21 and Figure 3.22. Based on this data days with one fill-draw cycle where the required exchange was not achieved, the filling cycle was short compared to the drawing cycle. This is also observed for days with two fill-draw cycles for the late fill-draw cycle.

$$\frac{\Delta V}{V} \approx \frac{\pi^{0.5} r_m d_1}{2V\sqrt{3}}$$

(Equation 3.9)

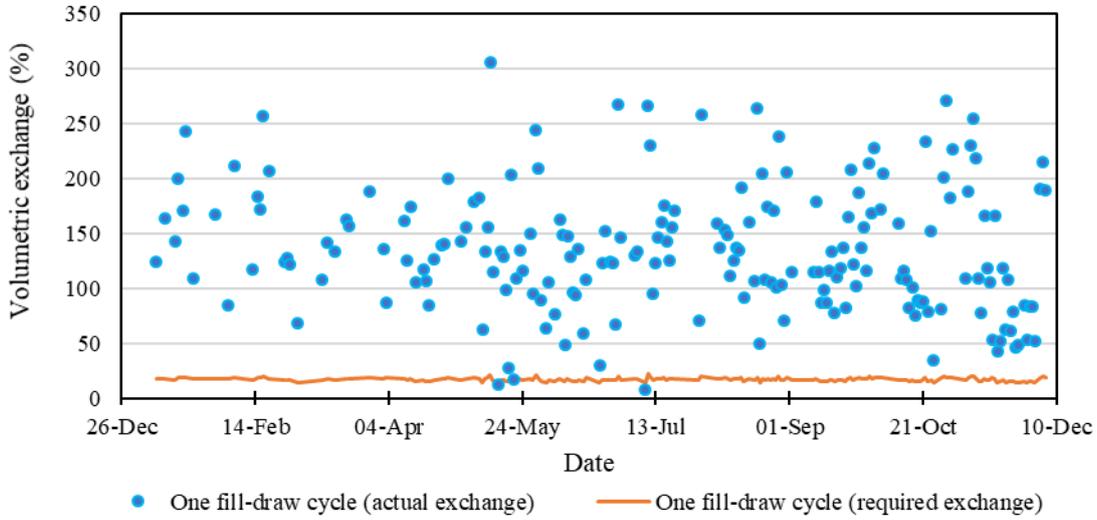


Figure 3.21: Volumetric exchange for days with one fill-draw cycle.

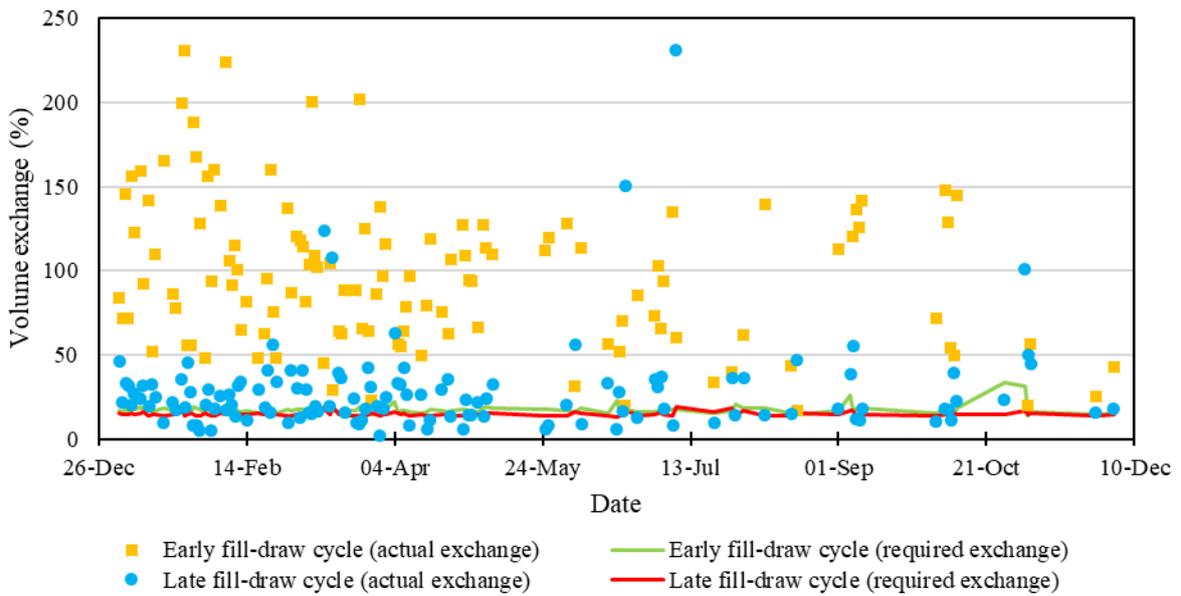


Figure 3.22: Volumetric exchange for days with two fill-draw cycles.

Possible Stagnation Zones

Temperature measurements throughout the reservoir was observed and the possible stagnation are shown in Table 3.7

Table 3.7: Temperature analysis and the potential link to stagnation

	<p>Early fill-draw cycle</p>
	<ul style="list-style-type: none"> • Possible stagnation in regions 6 and 15 • Good circulation in the middle region near the inlet and outlet (3, 5, 7, 9) • Region 10 is also a possible stagnant zone
	<p>Late fill-draw cycle</p>
	<ul style="list-style-type: none"> • Possible larger stagnation zone in regions 10, 12, 13 and 15

3.3.2 Profiling the microbial community in Reservoir A

3.3.2.1 Community membership/composition

The mean relative abundance (MRA) for all sample points and their repeats over the five different sampling months at a phylum level were evaluated (Figure 3.23). *Proteobacteria* and *Bacteroidetes* represented the highest abundances. Findings like these are not uncommon in drinking water systems. Specifically, *Proteobacteria* is more increased in August, September, October, November and December, while *Bacteroidetes* is higher in March; this variance could be accounted for by the time of year sampling was conducted, as temperatures differed based on seasonal change. Across all sampling points, there were minor differences in terms of their abundance from a month-to-month view. Variance in repeat samples within the same month shows no significant changes in abundance; this indicates that sequencing was successful, as replicates were reproducible. *Nitrospirae* and *Omnitrophicaeota* represent higher abundances in August than in other months. This is not uncommon as both these microbes are found in water environments. *Patescibacteria* was seen in lower abundances than in the other months, which is also known to be found in water environments, specifically freshwaters (Tian *et al.*, 2020).

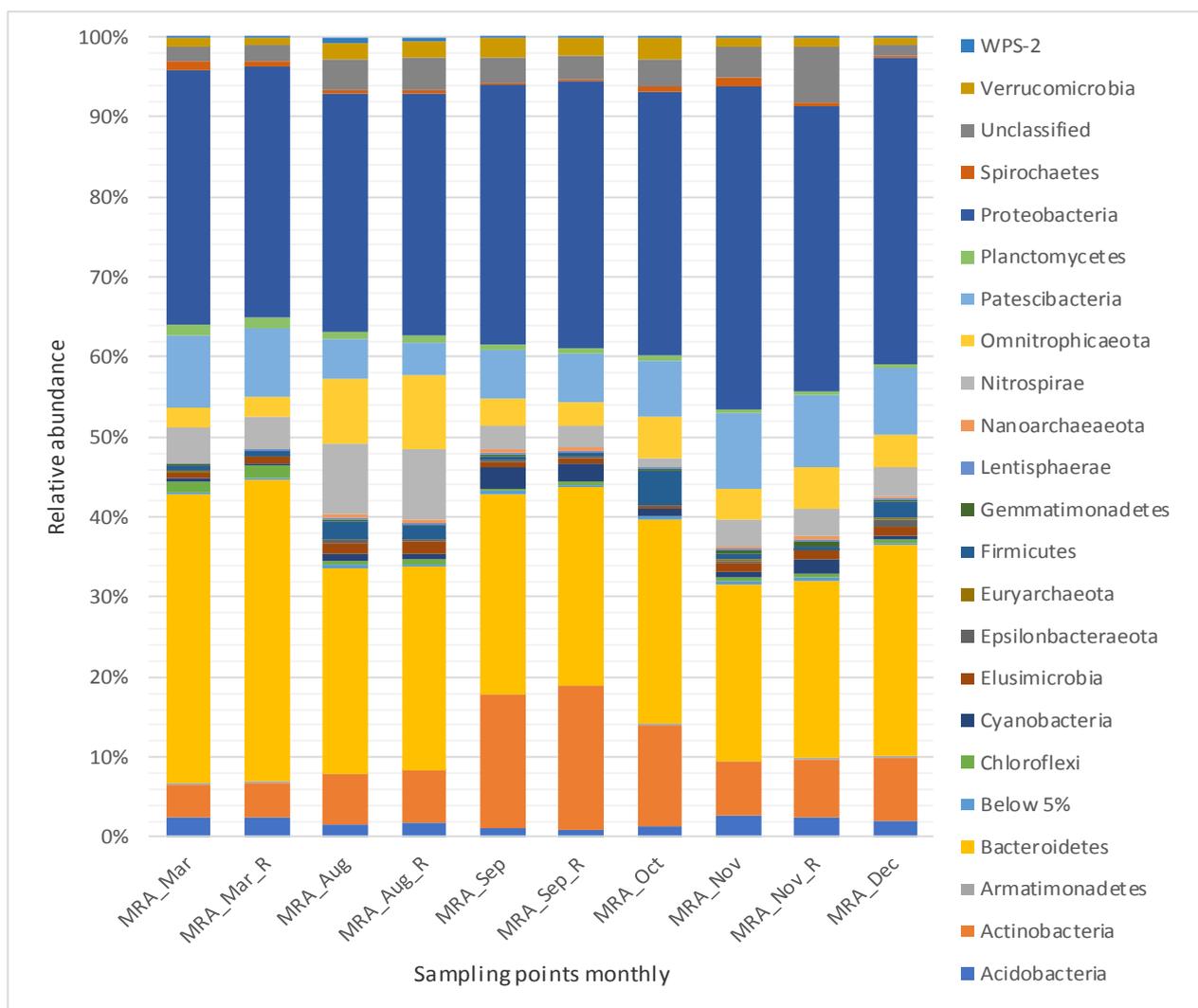


Figure 3.23: Representation of the mean relative abundance (MRA) of bacterial phyla of all sample points with their respective repeats, from month to month.

Actinobacteria show higher abundances in September and October months than other months. *Chloroflexi* showed higher abundances in March in comparison to other months; this could be due to March being a warmer month. A study done by Bennet *et al.*, 2020 showed that *Chloroflexi* has a distinct distribution based on temperature differences.

The mean relative abundance (MRA) of all sampling points was analysed by evaluating the communities at different depths within the reservoir at a phylum level (Figure 3.24). The microbial population looked very similar at different depths throughout the reservoir. Overall, *Proteobacteria* showed the highest relative abundances, with *Bacteroidetes* showing the second highest relative abundance. This indicates that a homogenous bacterial community is present in the reservoir, despite the different depths in the reservoir. This is important as it shows good microbial stability in the reservoir (Li *et al.*, 2019).

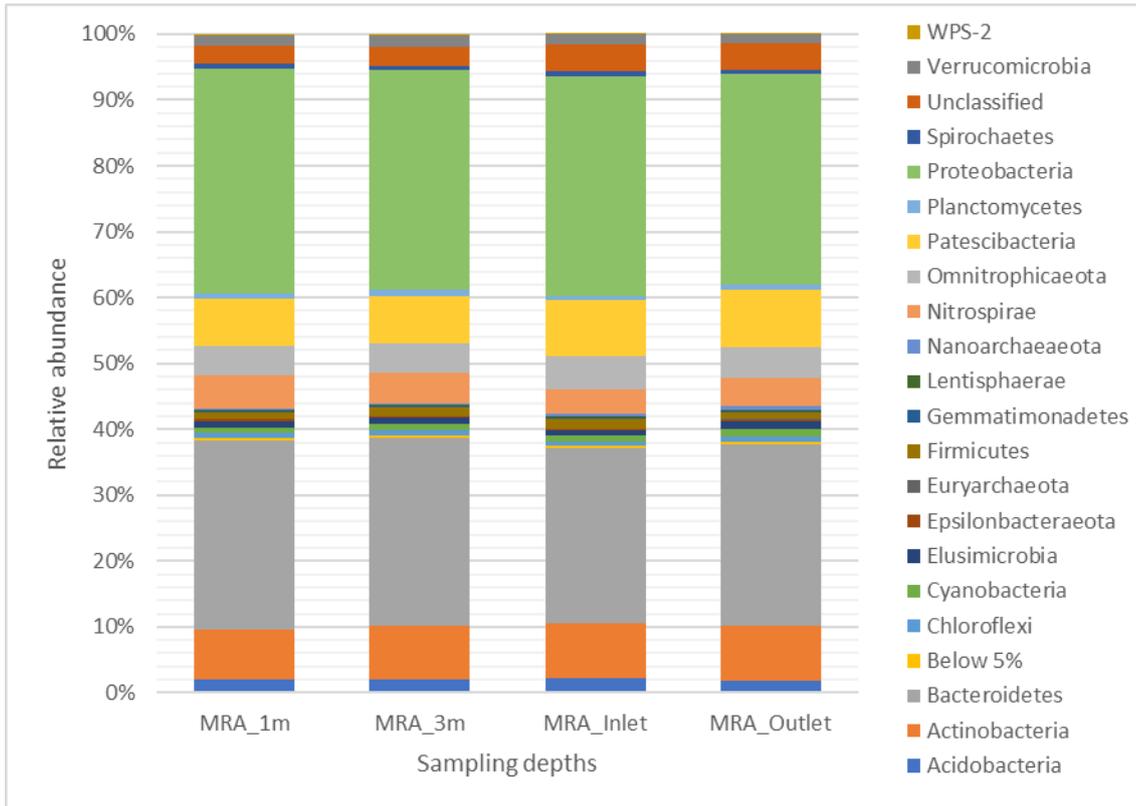


Figure 3.24: Representation of the mean relative abundances based on sample location (depths) over the 5 different months of sampling.

3.3.2.2 Alpha diversity

Alpha diversity measured showed three different measurement parameters. Species observed (Sobs), Shannon diversity index (H') and Pielou (J). Sobs is the number of species that are observed within each sample point and provides a good indication of richness in the system. H' examines diversity within the sample point by examining presence, absence and abundance. J examines evenness within the sample points by examining the numbers and abundance of species. Figure 3.25 is a representation of environmental parameters and their Shannon diversity, the higher the value, the richer in diversity a community is. These graphs show that there is high richness in diversity from month to month, from point to point, at different depths and within changing pH, free Cl and total Cl parameters.

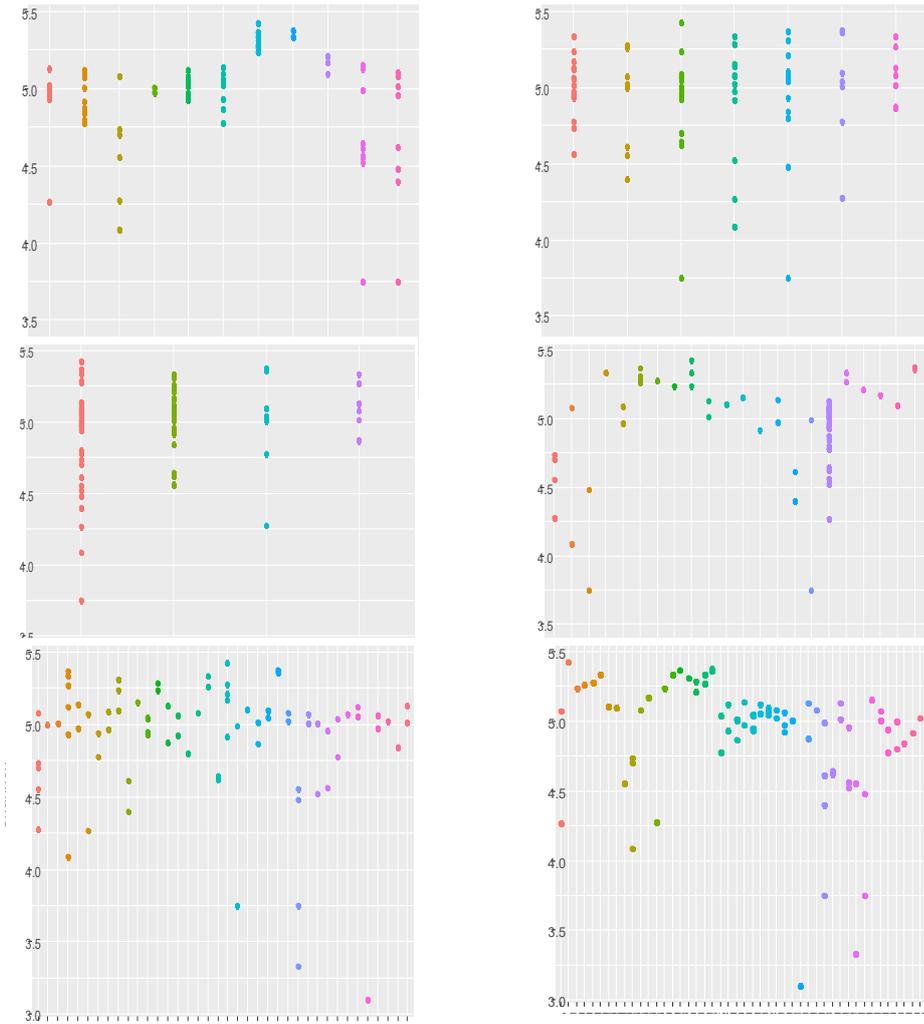


Figure 3.25: Shannon distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. pH ranged from 7 to 9. Higher values show increased richness in diversity.

Pielou evenness is represented by Figure 3.26 across all environmental parameters. Values closer to 100 show a higher evenness in diversity within the community. In our study the data shows the communities to have a high evenness with high consistency in the community. Sobs (Figure 3.27) showed a variable richness across all samples, the actual counts for Sobs varied from 50 to 600 and showed diversity in the numbers observed with all environmental parameters.

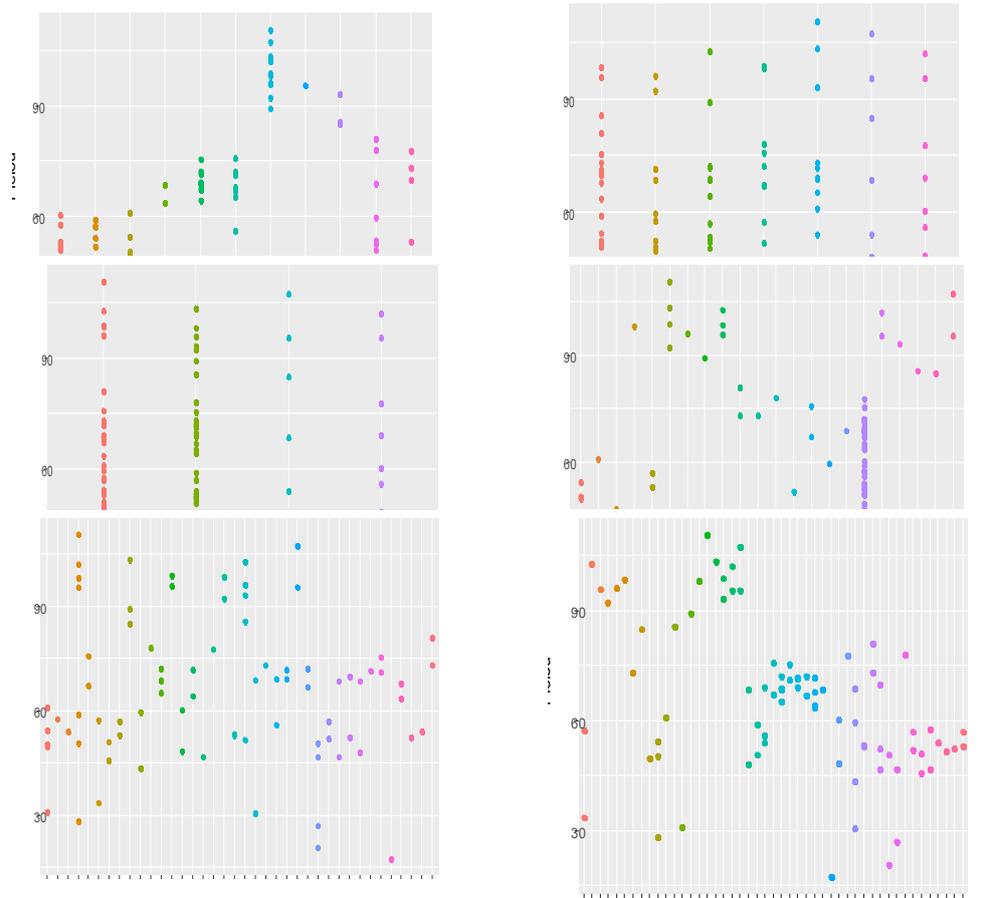


Figure 3.26: Pielou distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. pH ranged from 7 to 9. Values closer to 100 show a higher evenness distribution.

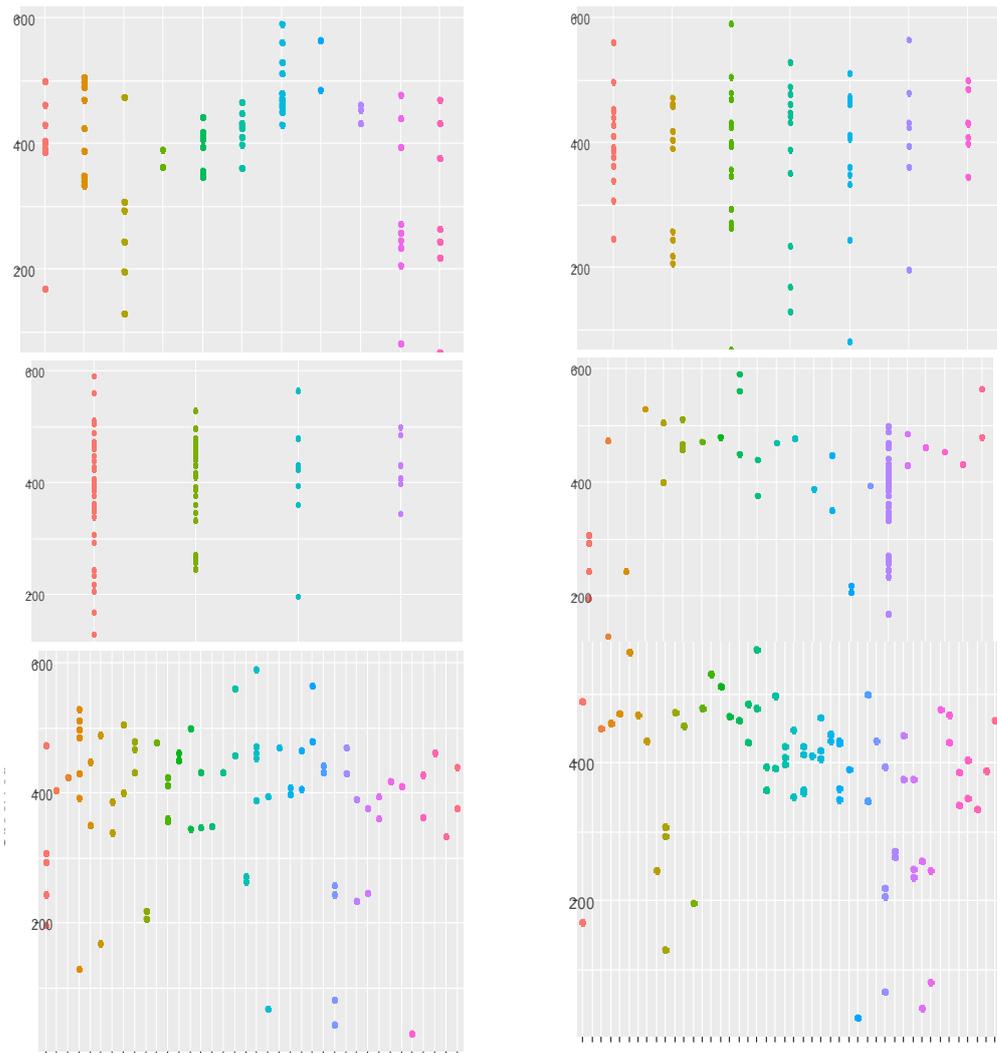


Figure 3.27: Observed distribution for all sample points in relation to months (top left), points location (top right), depth (middle left), total CI (middle right), free CI (bottom left) and pH (bottom right). Months included March, August, September, October, November, and December. Points location included sites 1, 2, 3, 4, 5, inlet and outlet. Depths included 1m, 3m, inlet and Outlet. Total CI and Free CI ranged from 1 to 2.2. Observed values range from 50 to 600, showing richness in diversity.

3.3.2.3 *Beta diversity observed*

Beta diversity examined the differences in bacterial composition among all sites by creating a distance matrix (a square matrix containing distances, this is a pairwise analysis between the two groups of interests, showing similarity or differences). Values closer to 1 indicate high dissimilarity among the community (Jaccard & Turrisi, 1990; Bray & Curtis, 1957). The closer the grouping are to one another the more closely they are related. Bray-Curtis determines community structure (incidence and abundance), while Jaccard determines community membership (absence and presence). Figures 3.28 and 3.29 shows sampling months and depth of all sample points, the groupings seem to cluster based on month, indicating a temporal change, while depth which is a spatial change, had little to no groupings.

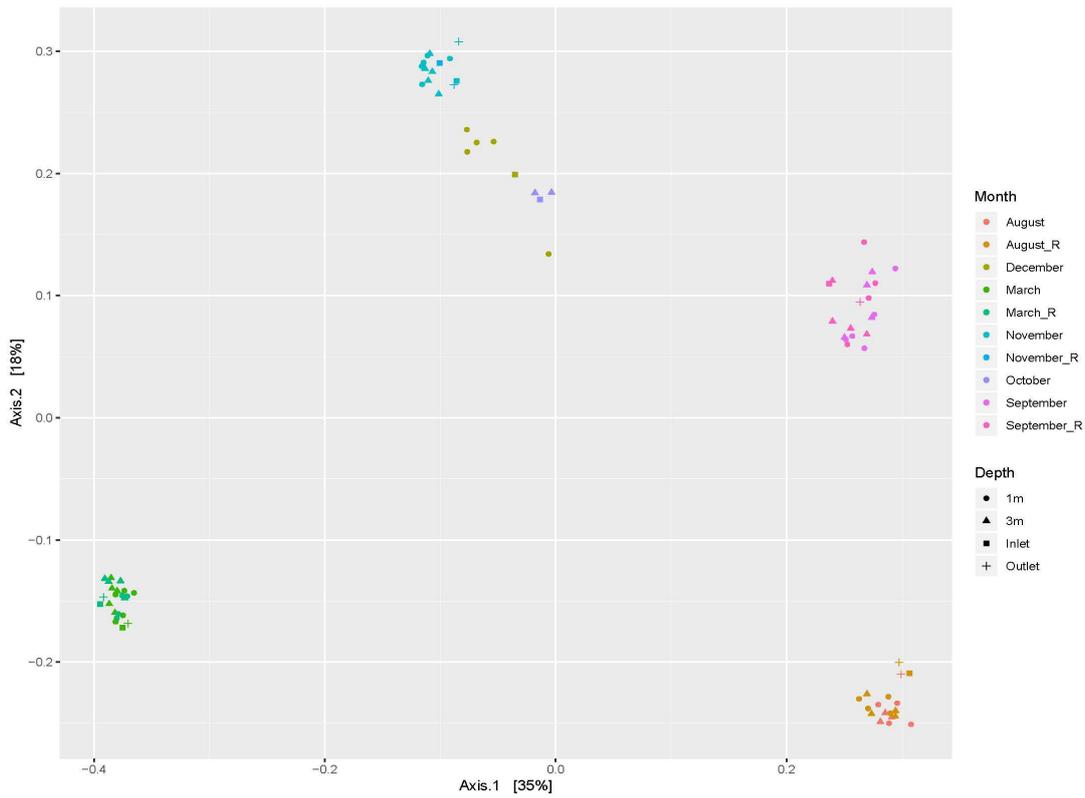


Figure 3.28: Principal coordination plot, showing groupings based on sampling months and depths using a Bray Curtis matrix.

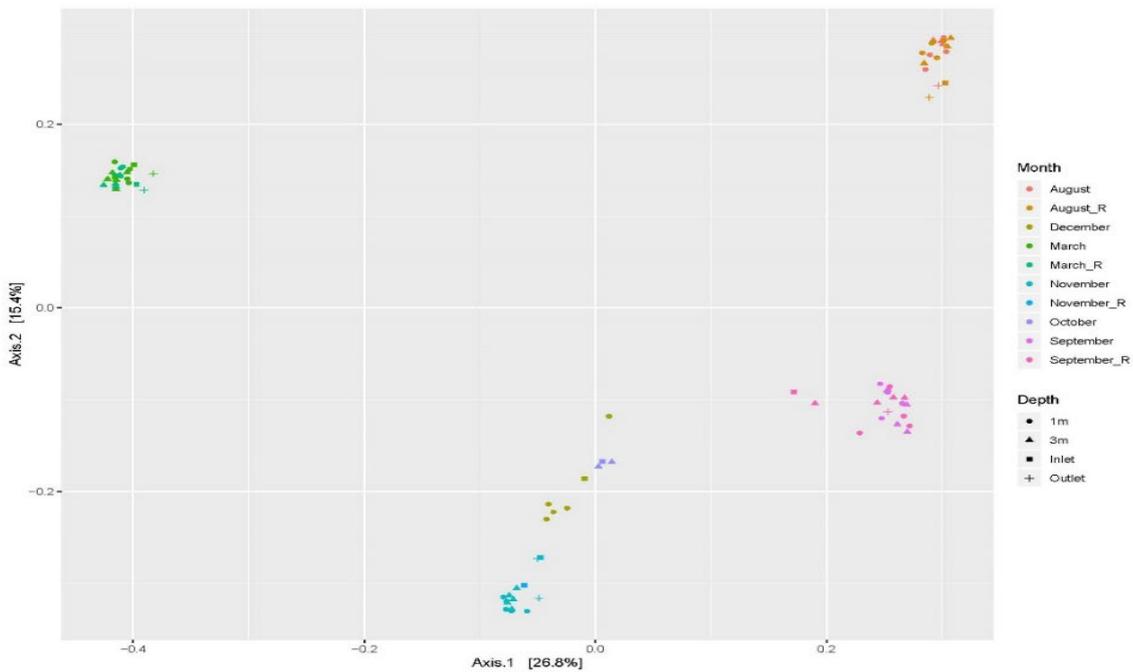


Figure 3.29: Principal coordination plot, showing groupings based on sampling months and depths using a Jaccard matrix.

3.3.3 Flow cytometry data for Reservoir A

Figure 3.30 represents flow cytometry cell counts, showing total, intact, and damaged cells. Cell counts were lower in August, October, November and December. March and September showed higher cell counts. Higher cell counts in March could be due to the warmer season. Warmer weather can increase microbial growth, especially in reservoirs as it's a closed environment and can thus act as an incubator. The increase in September was the highest cell count for all sampling months; the lack of changes in environmental conditions such as pH and chlorine led one to believe this could have been an isolated issue; perhaps the water quality was poorer than other months.

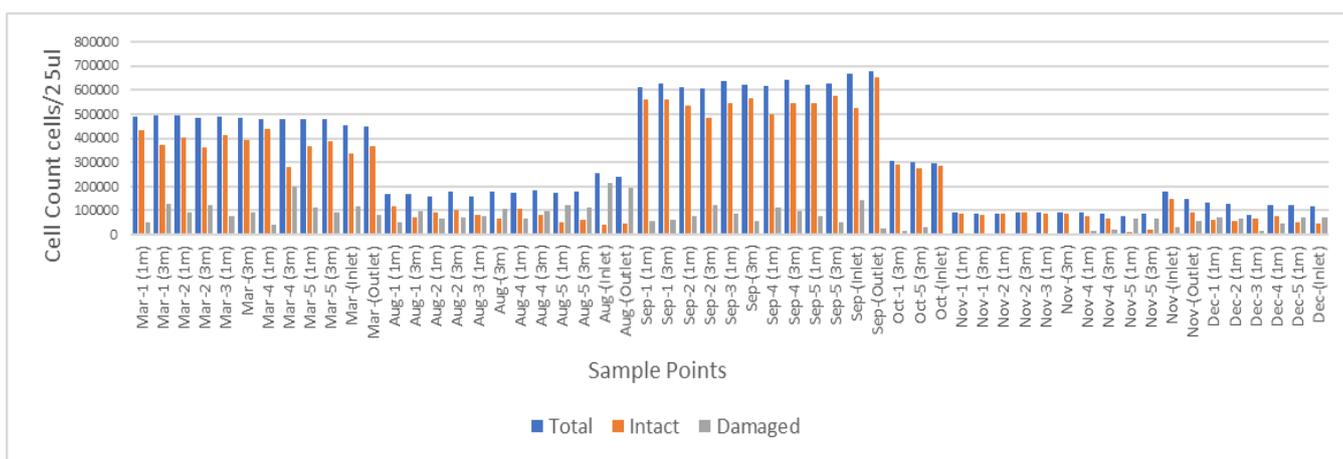


Figure 3.30: Graph showing total, intact and damaged flow cytometry cell counts.

Figure 3.31 shows a heatmap representing >1% of the total sequence abundance amplicon sequence variant (ASV-SEQ) are presented in Figure 3.31. The redder the blocks, the more abundant the ASV. A key for the horizontal axis of the heatmap forms part of Appendix B (Table B1-B2).

Bacteria seen in all sample points at high abundance included *Flavobacterium*, *Polynucleobacter*, *Burkholderiaceae*, *Porichthyaceae*, *Nitrospira* and *Sediminibacterium*. All these bacteria are commonly seen in drinking water systems. Genera such as *Flavobacterium* and *Nitrospira* are known to include autotrophic bacteria. Some samples had bacterial groups seen to be dominant that month than in other months for example, SEQ 33, 9 and 38 were seen in March at some sample points but little to none in the other sampling months. The heatmap gives a good understanding of the microbial community seen in the reservoir.

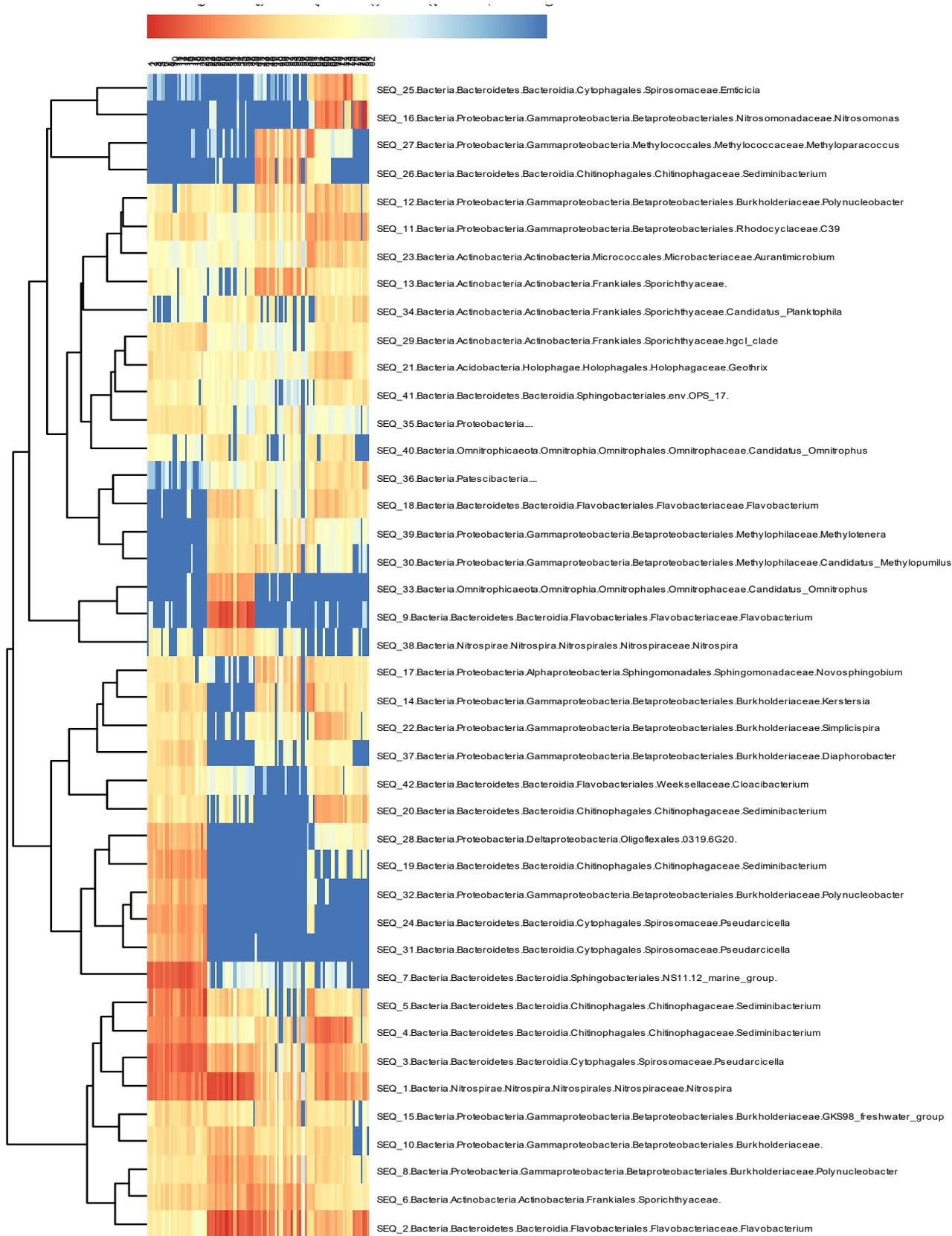


Figure 3.31: Representation of the abundance's levels of the 7 different sampling points over the different sampling months (Reservoir A). Heatmap shows most dominate sequences over the 7 sample points over the different months.

3.3.4 Functional role of dominant bacteria in a community Reservoir A

The analysis of the distance between samples within Reservoir A based on their Bray Curtis dissimilarities was examined (Figures 3.32-3.38). Spatial dynamics within the reservoir (distance between samples) and the change (Δ) in temperature between these points were also examined. However, although diversity can be seen within a sample, no significant differences were seen within a month.

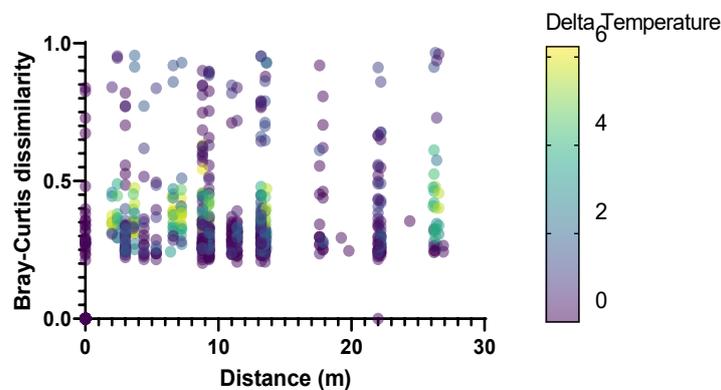


Figure 3.32: Distance (m) and change (Δ) temperature using pairwise structure-based Bray Curtis distances of all sample points.

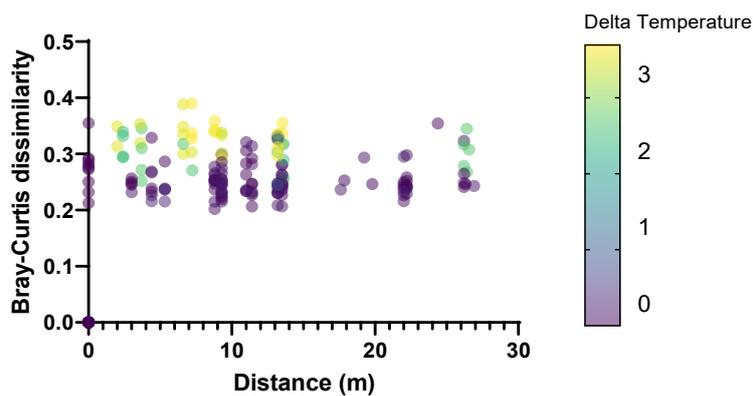


Figure 3.33: Distance (m) and change (Δ) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - March.

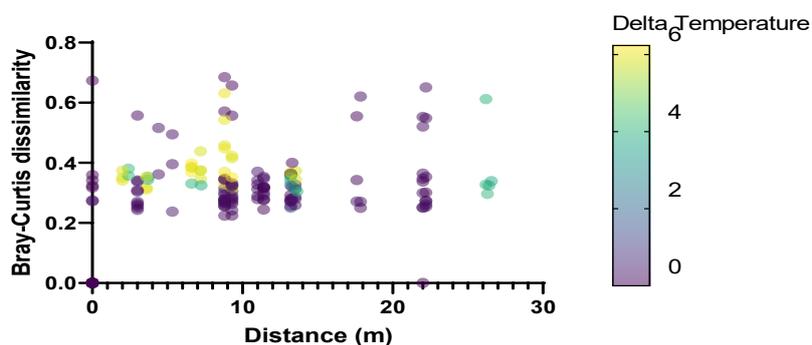


Figure 3.34: Distance (m) and change (Δ) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - August.

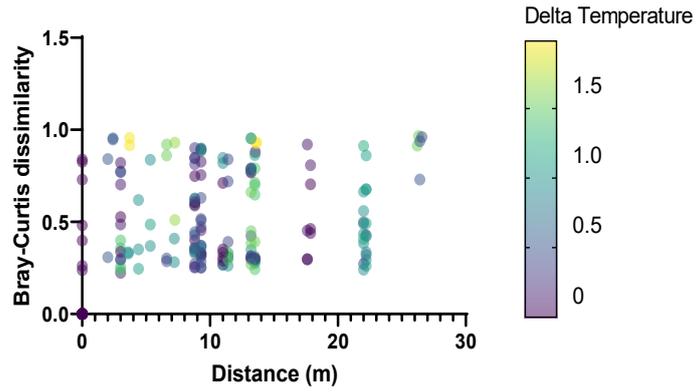


Figure 3.35: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - September.

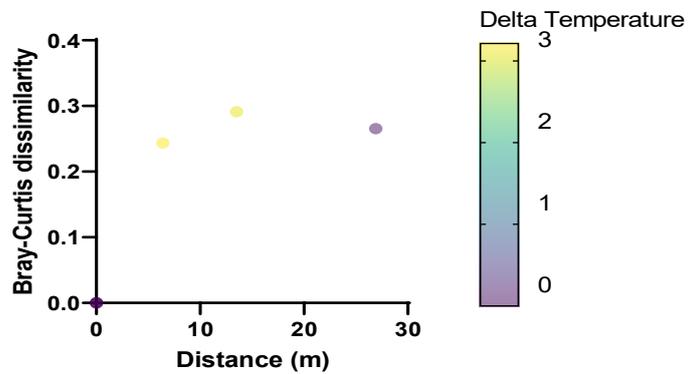


Figure 3.36: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - October.

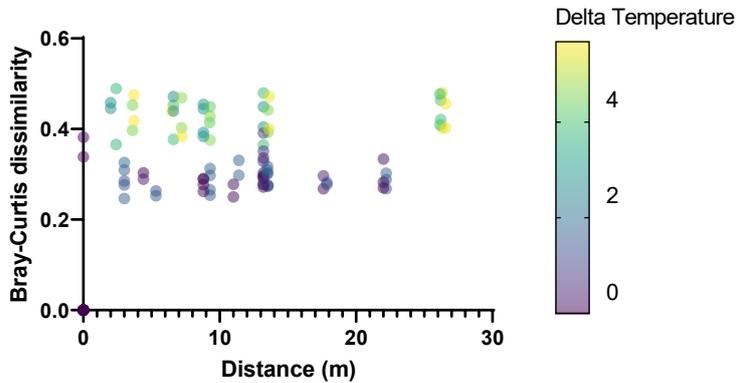


Figure 3.37: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - November.

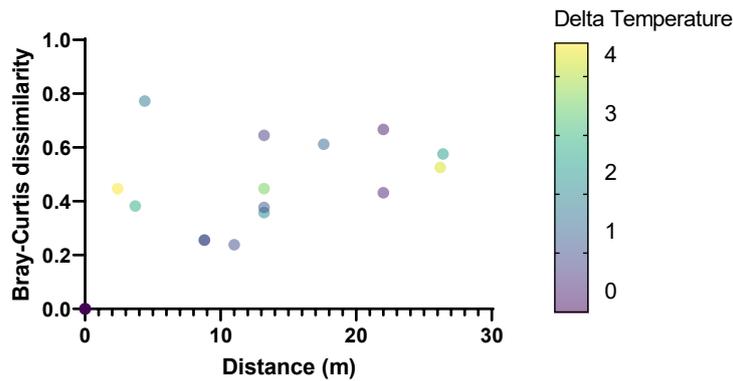


Figure 3.38: Distance (m) and change (delta) temperature using pairwise structure-based Bray Curtis distances of all sample points for the sampling month - December.

Although we observed higher temperature changes at shorter distances, there was no significant correlation between Bray Curtis and the distance or the temperature change within a sampling month. Samples were pooled and sequenced together, excluding October, due to a small sampling size (Table 3.8 represents the sample key).

Table 3.8: Key for pooled samples

Month	Key
March	Sample 1
August	Sample 2
September	Sample 3
November	Sample 4
December	Sample 5

From the co-assembly protocol in SqueezeMeta, bins kept were above 90% completeness and below 5% contamination; 999 co-assembly bins were produced, and 45 bins were kept (bins that fall within selected parameters). Table 3.9 represents the retained bins with their taxonomic classifications, which were established using SqueezeMeta, MiGA and GTDB-Tk databases. GTDB-Tk was classified to the fullest and was then selected to be reported. Bin classification was primarily assigned to *Proteobacteria* and *Bacteroidetes*.

Table 3.9: Representation of all retained bins with completeness above 90% and contamination below 5% contamination, including taxonomic classification of bins from the SqueezeMeta, GTDB-Tk and MiGA databases.

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
maxbin0152	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Sphingobacteriia	d: Bacteria p: Bacteroidota c: Bacteroidia o: AKYH767 f: B-17BO g: s:	100.00	3.50
metabat2133	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes	d: Bacteria p: Bacteroidota c: Bacteroidia o: NS11-12g f: g: s:	91.03	1.03
metabat21275	d: Bacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria c: Gammaproteobacteria o: UBA6186 f: UBA6186 g: UBA6186 s:	91.07	0.00
maxbin0017	d: Bacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Bdellovibrionota c: Bdellovibrionia_A o: UBA1018 f: UBA1018 g: s:	91.22	1.72
maxbin0219	d: Bacteria n_PVC group p: Candidatus Omnitrophica	d: Bacteria	d: Bacteria p: Omnitrophota c: Koll11 o: 2-02-FULL-51-18 f: 2-02-FULL-51-18 g: s:	91.38	0.00
metabat21278	d: Bacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Bdellovibrionota c: Bdellovibrionia_A o: UBA1018 f: g: s:	91.38	0.00
metabat2565	d: Bacteria	d: Bacteria	d: Bacteria p: Bdellovibrionota c: Oligoflexia o: Oligoflexales f: RGVZ01 g: s:	91.38	0.00
metabat2825	d: Bacteria n_PVC group p: Candidatus Omnitrophica	d: Bacteria	d: Bacteria p: Omnitrophota c: Omnitrophia o: Omnitrophales f: UBA2337 g: UBA2337 s:	91.38	0.00

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
metabat21083	d: Bacteria n_Terrabacteria group p: Firmicutes c: Bacilli o: Bacillales f: Bacillaceae g: Fictibacillus	d: Bacteria	d: Bacteria p: Firmicutes c: Bacilli o: Bacillales_G f: Fictibacillaceae g: Fictibacillus s:	91.47	2.10
metabat2871	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Bdellovibrionota c: Bdellovibrionia_A o: UBA1018 f: UBA1018 g: s:	91.54	0.00
metabat2805	d: Bacteria p: Proteobacteria c: Betaproteobacteria n_unclassified Betaproteobacteria s: Betaproteobacteria bacterium	d: Bacteria p: Proteobacteria c: Betaproteobac teria	d: Bacteria p: Proteobacteria c: Gammaproteobac teria o: Burkholderiales f: Burkholderiaceae g: PHCI01 s: PHCI01 sp002842205	91.86	4.84
metabat2697	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria c: Deltaproteoba cteria	d: Bacteria p: Proteobacteria c: Alphaproteobacte ria o: Rickettsiales f: SXRF01 g: s:	92.33	2.64
metabat21372	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria c: Alphaproteobacte ria o: Rickettsiales f: SXRF01 g: s:	92.86	0.00
maxbin0625	d: Bacteria	d: Bacteria	d: Bacteria p: Patescibacteria c: Gracilibacteria o: UBA1369 f: UBA1369 g: PALSA-1335 s:	92.95	0.31
metabat21391	d: Bacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Patescibacteria c: Gracilibacteria o: UM-FILTER-43- 11 f: g: s:	93.10	0.00
metabat2340	d: Bacteria n_PVC group p: Candidatus Omnitrophica	d: Bacteria	d: Bacteria p: Omnitrophota c: Omnitrophia o: Omnitrophales f: GWA2-52-8 g: s:	93.10	1.72

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
metabat2533	d: Bacteria	d: Bacteria	d: Bacteria p: Elusimicrobiota c: Elusimicrobia o: F11 f: F11 g: s:	93.10	0.00
metabat292	d: Bacteria n_PVC group p: Candidatus Omnitrophica	d: Bacteria	d: Bacteria p: Omnitrophota c: Omnitrophia o: Omnitrophales f: GWA2-52-8 g: s:	93.10	0.00
metabat2935	d: Bacteria	d: Bacteria	d: Bacteria p: Myxococcota c: XYA12-FULL-58- 9 o: f: g: s:	93.10	1.72
metabat21132	d: Bacteria n_FCB group n_Bacteroidetes/Ch lorobi group p: Bacteroidetes c: Chitinophagia o: Chitinophagales	d: Bacteria p: Bacteroidetes c: Chitinophagia	d: Bacteria p: Bacteroidota c: Bacteroidia o: Chitinophagales f: Chitinophagaceae g: Ferruginibacter s:	93.59	0.05
metabat2136	d: Bacteria n_FCB group n_Bacteroidetes/Ch lorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Sphingobacteri ia	d: Bacteria p: Bacteroidota c: Bacteroidia o: AKYH767 f: B- 17BO g: UBA2475 s:	93.59	4.87
metabat2954	d: Bacteria p: Proteobacteria c: Alphaproteobacteri a o: Sphingomonadales f: Sphingomonadace ae	d: Bacteria	d: Bacteria p: Proteobacteria c: Alphaproteobacte ria o: Sphingomonadale s f: Sphingomonadac eae g: Novosphingobium s:	93.72	1.77
maxbin0237	d: Bacteria n_FCB group n_Bacteroidetes/Ch lorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes	d: Bacteria p: Bacteroidota c: Bacteroidia o: NS11-12g f: UBA955 g: VMCP01 s:	93.89	1.03
metabat2230	d: Bacteria	d: Bacteria	d: Bacteria p: Elusimicrobiota c: Elusimicrobia o: F11 f: F11 g: s:	93.97	3.45
metabat21159	d: Bacteria p: Proteobacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Bdellovibrionota	94.12	1.68

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
		c: Deltaproteobacteria	c: Bdellovibrionia o: Bdellovibrionales f: Bdellovibrionaceae g: s:		
metabat2404	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes c: Cytophagia o: Cytophagales f: Cytophagaceae g: Runella	d: Bacteria p: Bacteroidetes c: Cytophagia	d: Bacteria p: Bacteroidota c: Bacteroidia o: Cytophagales f: Spirosomaceae g: Runella s:	94.65	4.61
metabat2300	d: Bacteria	d: Bacteria	d: Bacteria p: Bdellovibrionota c: UBA2428 o: UBA2428 f: UBA2428 g: s:	94.67	0.00
metabat21314	d: Bacteria	d: Bacteria	d: Bacteria p: Elusimicrobiota c: Elusimicrobia o: F11 f: F11 g: s:	94.83	1.72
metabat21373	d: Bacteria	d: Bacteria	d: Bacteria p: Planctomycetota c: MHYJ01 o: WSZJ01 f: g: s:	94.83	3.45
metabat21381	d: Bacteria n_PVC group p: Candidatus Omnitrophica	d: Bacteria	d: Bacteria p: Omnitrophota c: Koll11 o: UBA10015 f: g: s:	94.83	1.72
metabat21422	d: Bacteria	d: Bacteria	d: Bacteria p: Margulisbacteria c: Riflemargulisbacteria o: GWF2-35-9 f: GWF2-35-9 g: s:	94.83	3.45
metabat21339	d: Bacteria p: Proteobacteria c: Betaproteobacteria	d: Bacteria p: Proteobacteria c: Betaproteobacteria	d: Bacteria p: Proteobacteria c: Gammaproteobacteria o: Burkholderiales f: UKL13-2 g: s:	95.27	0.96
maxbin1629	d: Bacteria n_Terrabacteria	d: Bacteria p: Actinobacteria	d: Bacteria p: Actinobacteriota	95.52	4.27

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
	group p: Actinobacteria c: Actinomycetia		c: Actinomycetia o: Actinomycetales f: Microbacteriaceae g: Aquiluna s:		
metabat21071	d: Bacteria	d: Bacteria	d: Bacteria p: Spirochaetota c: o: f: g: s:	95.69	1.72
metabat21347	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Chitinophagia	d: Bacteria p: Bacteroidota c: Bacteroidia o: Chitinophagales f: Chitinophagaceae g: UBA1312 s:	96.15	1.79
metabat2573	d: Bacteria	d: Bacteria p: Chlorobi	d: Bacteria p: Bacteroidota c: UBA10030 o: UBA10030 f: UBA6906 g: CAADGV01 s	96.24	3.45
metabat2983	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Sphingobacteri ia	d: Bacteria p: Bacteroidota c: Bacteroidia o: AKYH767 f: 2-12- FULL-35-15 g: s:	96.38	2.56
maxbin0200	d: Bacteria	d: Bacteria	d: Bacteria p: Bdellovibrionota c: Bacteriovoracia o: Bacteriovoracales f: Bacteriovoracace ae g: s:	96.55	2.54
metabat21080	d: Bacteria	d: Bacteria p: Proteobacteria	d: Bacteria p: Bdellovibrionota c: Bdellovibrionia_A o: UBA1018 f: g: s:	96.55	2.07
metabat240	d: Bacteria n_Terrabacteria group p: Actinobacteria c: Actinomycetia o: Corynebacteriales f: Mycobacteriaceae g: Mycolicibacter	d: Bacteria p: Actinobacteria c: Actinobacteria o: Corynebacteri ales	d: Bacteria p: Actinobacteriota c: Actinomycetia o: Mycobacteriales f: Mycobacteriaceae g: Mycobacterium s: Mycobacterium arupense	96.79	2.48

Bin ID	Taxonomy SqueezeMeta	Taxonomy MiGA	Taxonomy GTDB-Tk	Completeness	Contamination
metabat21496	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Sphingobacteriia	d: Bacteria p: Bacteroidota c: Bacteroidia o: AKYH767-A f: JABDAW01 g: s:	96.92	0.00
metabat21078	d: Bacteria	d: Bacteria p: Bacteroidetes	d: Bacteria p: Bacteroidota c: Kapabacteria o: Kapabacteriales f: UBA961 g: s:	98.28	0.00
maxbin0414	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes c: Sphingobacteriia	d: Bacteria p: Bacteroidota c: Bacteroidia o: Sphingobacteriales f: Sphingobacteriaceae g: Daejeonella s:	98.46	4.63
metabat27	d: Bacteria n_FCB group n_Bacteroidetes/Chlorobi group p: Bacteroidetes	d: Bacteria p: Bacteroidetes	d: Bacteria p: Bacteroidota c: Bacteroidia o: NS11-12g f: UKL13-3 g: s:	99.23	0.00

Transcripts per million (TPM) were drawn from the SqueezeMeta results. These values indicate the number of times the feature is observed when randomly sampling 1 million features, i.e. the abundances of the different bins within the sample. The values can be observed in Figure 3.39 (TPM values) and Figure 3.40 (abundance values). Sample 3 showed to be the most variable in comparison to other samples. High abundances of metabat 2404 (d: *Bacteria* p: *Bacteroidota* c: *Bacteroidia* o: *Cytophagales* f: *Spirosomaceae* g: *Runella*) and maxbin 1629 (d: *Bacteria* p: *Actinobacteriota* c: *Actinomycetia* o: *Actinomycetales* f: *Microbacteriaceae* g: *Aquiluna*) were seen. Sample 2 showed high abundances of metabat 21275 (d: *Bacteria* p: *Proteobacteria* c: *Gammaproteobacteria* o: UBA6186 f: UBA6186 g: UBA6186). High bin abundances were seen in sample 1, specifically for maxbin 0017(d: *Bacteria* p: *Bdellovibrionota* c: *Bdellovibrionia_A* o: UBA1018 f: UBA1018), maxbin 0219 (d: *Bacteria* p: *Omnitrophota* c: Koll11 o: 2-02-FULL-51-18 f: 2-02-FULL-51-18), metabat 21422 (d: *Bacteria* p: *Margulisbacteria* c: *Riflemargulisbacteria* o: GWF2-35-9 f: GWF2-35-9), metabat 2573 (d: *Bacteria* p: *Bacteroidota* c: UBA10030 o: UBA10030 f: UBA6906 g: CAADGV01), metabat 2983 (d: *Bacteria* p: *Bacteroidota* c: Bacteroidia o: AKYH767 f: 2-12-FULL-35-15), maxbin 0200 (d: *Bacteria* p: *Bdellovibrionota* c: *Bacteriovoracia* o: *Bacteriovoracales* f: *Bacteriovoracaceae*) and metabat 27 (d: *Bacteria* p: *Bacteroidota* c: *Bacteroidia* o: NS11-12g f: UKL13-3).

KEGG decoder and KEGG MODULE (based on KO IDs) allowed for functional interpretation. KEGG divides pathways based on modules. The KEGG database manually curates a collection of modular functional units, categorized into pathways, signature, and reaction modules, allowing for functional annotation of microorganisms. For a pathway to be complete, the module numbers associated with that pathway need to be complete.

Figure 3.41 shows a representation of KEGG-associated KO IDs with the different bins (the redder the clocks, the more complete the pathway). A few pathways were mostly complete among the bins, such as glycolysis, glucogenesis and the TCA cycle, to name a few.

A summary of the pathways associated with each sample is shown in Figure 3.42. It can be observed that when looking at each sample, most pathways were complete, showing the presence of functionality. However, there were a few pathways that showed variance. For example, alt thiosulphate oxidation *tsdA* (absent in sample 3 only), DMSP demethylation (present in sample 5 only), alcohol oxidase (present in sample 4 and 5), adhesion (present in sample 1 and 5). Although most functional pathway were almost all present for all months, it is worth mentioning functional pathways that were completely absent from all samples. They include pectin esterase, exo-poly-alpha-galacturonosidase, bifunctional chitinase/lysozyme, hydrazine dehydrogenase, hydrazine synthase, sulphur reductase *sreABC*, sulphur disproportionation, DMSP demethylation, NiFe hydrogenase, methanogenesis, biofilm regulator, colonic acid and biofilm protein A, curli fimbriae, cobalt transporter *CtbtA* (Table 3.10).



Figure 3.39: Represents transcripts per million for each bin against their sample points.

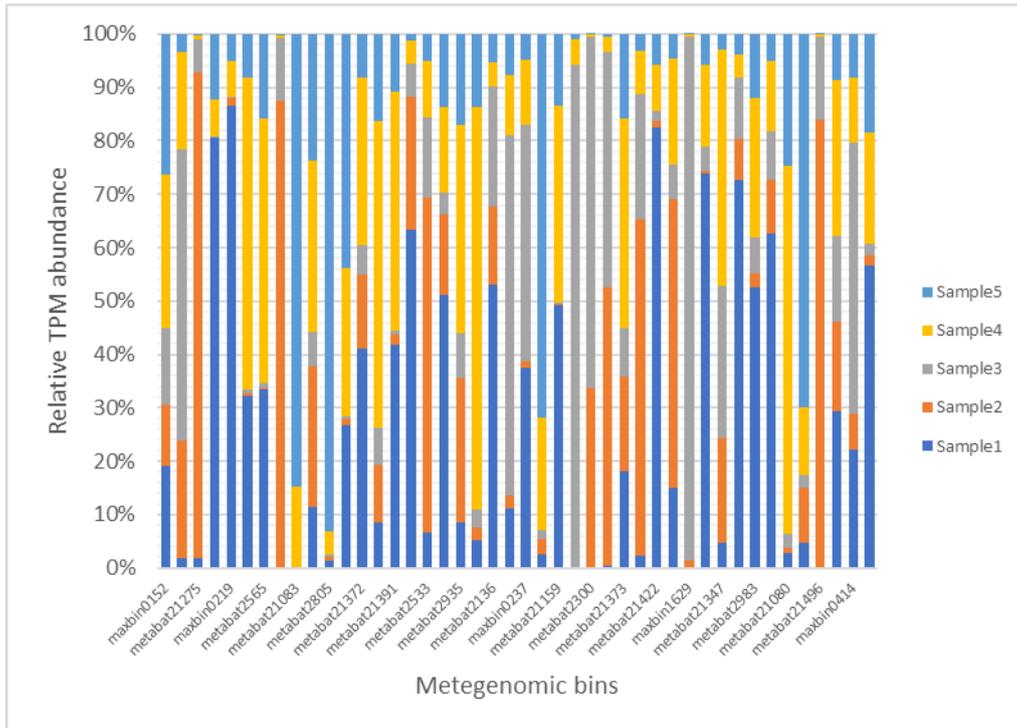


Figure 3.40: Represents relative transcripts per million abundances of metagenomic bins.

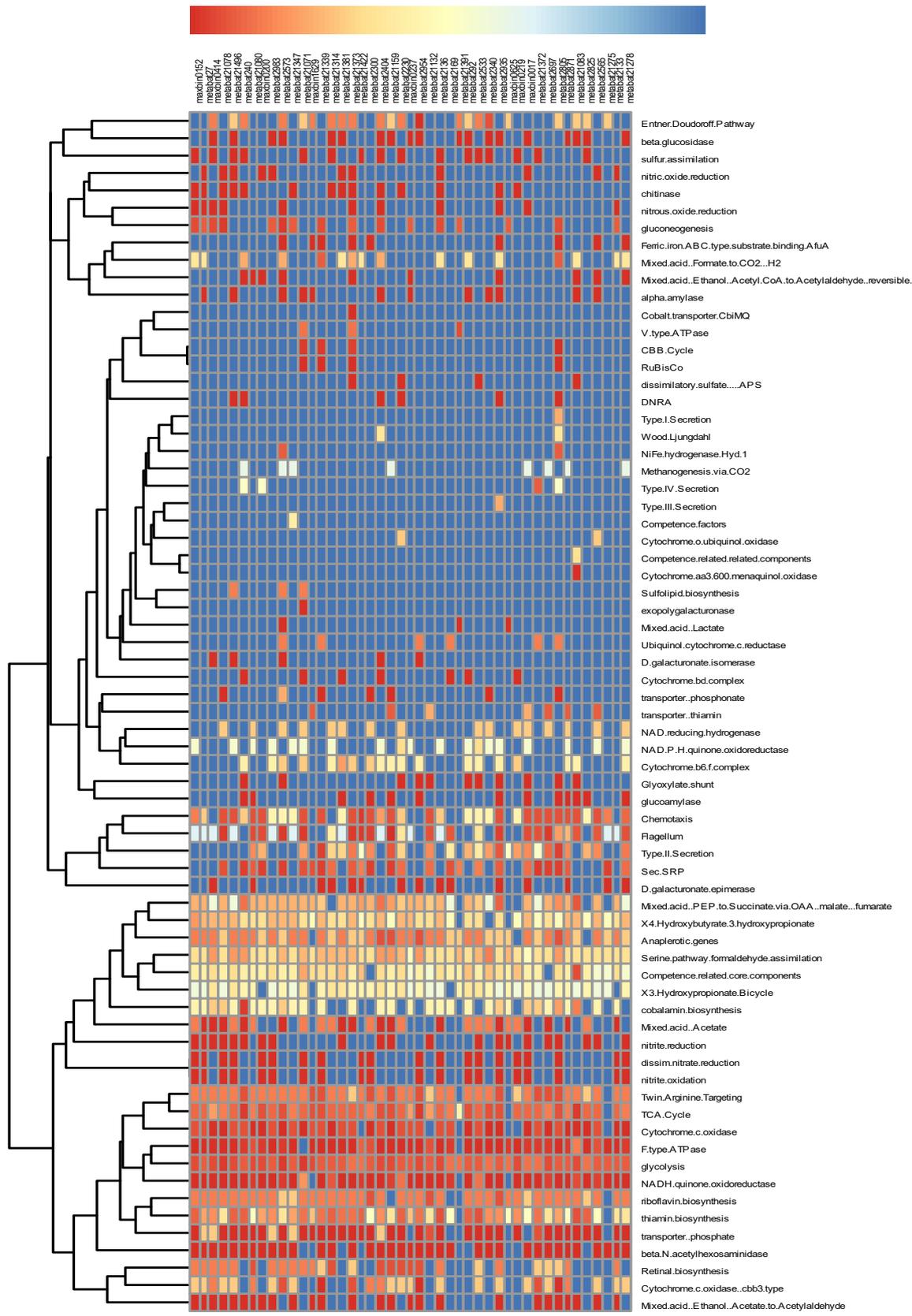


Figure 3.41: Representation of KEGG pathways associated with the 45 high quality bins obtained from the study (the redder the clocks the more complete the pathway, pathways missing from all bins are excluded).

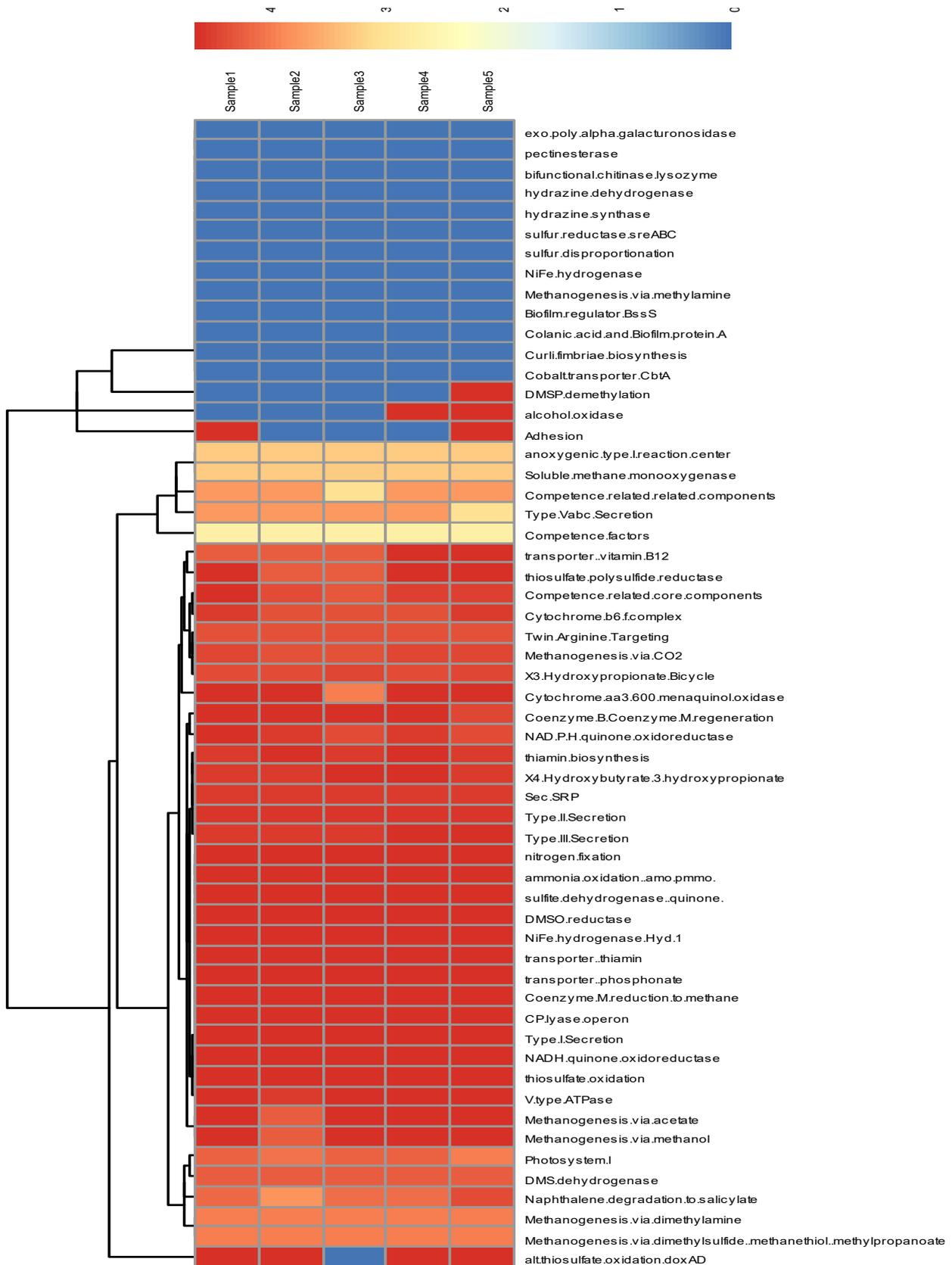


Figure 3.42: Summary of pathways associated with each sample which are representative of the bacterial community in Reservoir A at the time of sampling (the redder the clocks the more complete the pathway, complete pathways found in all communities are excluded).

Table 3.10: Unique functional abilities either present or absent in some of the samples

Function	Mechanism	Unique	Ability
Thiosulphate oxidation <i>tsdA</i>	Sulphur metabolism	Absent sample 1	Converting thiosulfate to tetrathionate (Brito <i>et al.</i> , 2015).
DMSP demethylation	Methane metabolism	Present sample 2	The DMSP demethylation pathway consists of a series of reactions that convert DMSP into methanethiol (MeSH), HS-CoA, CO ₂ , and acetaldehyde (Reisch <i>et al.</i> , 2011).
Alcohol oxidase	Fatty acid metabolism	Present samples 4 & 5	C-type haemoprotein and oxidation of alcohols (Cheng <i>et al.</i> , 2005).
Adhesion	Signalling molecules and interaction	Present samples 1 & 5	Bundles of actin filaments are anchored to allow for attachment.
Pectin esterase	Cell wall enzyme	Absent all samples	Pectin esterase catalyses the de-esterification of pectin into pectate and methanol.
Exo-poly-alpha-galacturonosidase,	Enzyme	Absent all samples	Alpha-galactosidase A breaks down a molecule called globotriaosylceramide, which consists of three sugars attached to a fatty substance.
Bifunctional chitinase/lysozyme,	Enzyme	Absent all samples	The enzyme binds to chitin and randomly cleaves glycosidic linkages in chitin and chitodextrins in a non-processive mode.
Esterase hydrazine dehydrogenase,	Enzyme	Absent all samples	The enzyme, which is involved in the pathway of anaerobic ammonium oxidation in anammox bacteria (Kanehisa & Goto, 2000).
Hydrazine synthase	Enzyme	Absent all samples	The enzyme, characterized from anaerobic ammonia oxidizers (anammox bacteria), is one of only a few enzymes that are known to form an N-N bond (Kanehisa & Goto, 2000).
Sulphur reductase <i>sreABC</i> ,	Enzyme	Absent all samples	An iron-sulfur protein. The enzyme from the hyperthermophilic archaeon <i>Pyrococcus furiosus</i> is part of two heterotetrameric complexes where the beta and gamma subunits function as sulfur reductase and the alpha and delta subunits function as hydrogenases (Kanehisa & Goto, 2000).
Sulphur disproportionation	Sulphur metabolism	Absent all samples	Inorganic fermentation as one sulfur compound serves as electron donor and acceptor (Kanehisa & Goto, 2000).
NiFe hydrogenase	Enzyme	Absent all samples	An iron-sulfur protein. Some forms of the enzyme contain nickel (Kanehisa & Goto, 2000).
Methanogenesis	Methane metabolism	Absent all samples	Methanogenesis is an anaerobic respiration that generates methane as the final product of metabolism (Kanehisa & Goto, 2000).
Biofilm regulator	Enzyme	Absent all samples	Allows for signalling (Kanehisa & Goto, 2000).

Function	Mechanism	Unique	Ability
Colanic acid and biofilm protein A	Protein	Absent all samples	Biofilm transcription regulator (Kanehisa & Goto, 2000).
Curli fimbriae	Protein	Absent all samples	Curli fimbria is a fibrous surface protein that is important for biofilm development by E (Kanehisa & Goto, 2000).
Cobalt transporter CtbA	Protein	Absent all samples	Involved in transportation.

3.4 SUMMARY

The assessment of the design of Reservoir A showed that depending on the fill-draw cycle regions of stagnation could be predicted. It is believed that a late fill-draw cycle could have a larger stagnation zone directly opposite the inlet on the other side of the reservoir. While there may have been a concern for drinking water system reservoirs due to increased microbial growth, this study has shown that the microbiology is rather stable. The 16S analysis showed no significant changes within the reservoir, but only from month to month. For this reason, DNA extracted was pooled based on months and sequenced for metagenomic sequencing.

The sequences showed Proteobacteria and Bacteroidetes as the most abundant in the community. Deep taxonomy classification proved to be demanding of the bins compiled, indicating there could be novel species within the community. The only bin that could be identified to species level was Metabat240 which represented *Mycobacterium arupense*. This species is known as an opportunistic pathogen.

CHAPTER 4: THE EFFECT OF RESIDENCE TIME ON THE MICROBIAL WATER COMMUNITY IN RESERVOIRS

4.1 INTRODUCTION

Most South African distribution networks supply water to different communities, sometimes hundreds of kilometres from the source water. To ensure even supply throughout the day, water is typically stored in large service or community reservoirs from where it is pumped to different communities (Prest *et al.*, 2016). As service reservoirs store water for periods of time the microbial quality can become questionable (Fard and Barkdoll, 2018). The reduction in water quality in these reservoirs is becoming an increasing concern for municipalities responsible for supplying clean drinking water (Rossman and Grayman, 1999; LeChevallier *et al.*, 1996; Marek *et al.*, 2007; Fard and Barkdoll, 2018). Reservoirs is a critical component of distribution networks and it is therefore important to understand their microbial ecology in greater detail.

The approach to designing large reservoirs is very conservative, and it can consequently reduce the water quality (Basile *et al.*, 2008; van Zyl and Haarhoff, 2007). Because of the extensive nature of service reservoirs, it is nearly impossible to have a homogeneously mixed system (Zhang *et al.*, 2013). Stagnated zones, often called dead zones, describe areas with slow-flowing water or water that does not flow. These stagnated zones can cause adverse public health effects due to increased microbial growth in the system (Fard and Barkdoll, 2018). Observing and understanding the water quality in reservoirs would require sampling from different zones and depths.

A previous study showed increased levels of microbes in these reservoirs. Identifying the microorganisms found in reservoirs and establishing their functions could aid in managing reservoirs. All the interactions between microbes and their environment and conditions in the reservoir could be explained by understanding and investigating possible functional networks, which would explain why microbes survive and can increase. In Chapter 6 we demonstrated that 16S profiling revealed a consistent homogeneous community throughout the reservoir on any specific sampling date, with high diversity, regardless of depth. For this part of the study a metagenomic study was performed to get a better understanding of the functional potential of the microbes in the system.

In this portion of the study, permission was obtained a local municipality in Gauteng to include one of their community reservoir in this study. Reservoir B receives treated drinking water from a large water treatment works. This reservoir was investigated to determine the effect of residence time on the microbial water community in the reservoir. After the filling of the reservoir, the inlet to the reservoir was closed and not refilled until the reservoir dropped to a level of 35%. This was done to allow for the longest possible residence time in the system. Sampling was conducted over a period of a week. Samples were processed and analysed as indicated above.

4.2 METHODS

4.2.1 Reservoir description

Sampling was done at an additional community reservoir (Reservoir B) to determine the effect of residence time on the microbial water community in the reservoir over time. This reservoir received

treated drinking water from a large water treatment works. After the filling of the reservoir, the inlet to the reservoir was closed until the reservoir reached a level of 35%. This was done to allow for the longest possible residence time in the system.

4.2.2 Sample collection and analysis

Sampling was conducted over a period of a week. Sampling was conducted 3 times a day (early, midday and late) and 3 different sampling points (Source inlet, Reservoir outlet and Tower outlet). Samples were sampled in duplicates, and some were sent for sequencing. Approximately 8 L of each sample was collected in sterile 8 L Large Narrow Mouth Nalgene polycarbonate bottles (Thermo Scientific™, South Africa) and transported on ice to the laboratory. and stored in a large refrigerator. The samples were concentrated using STERIVEX™-GP 0.22 µm polycarbonate membrane filter unit (Merck Millipore, South Africa). Filtration was facilitated using a Watson-Marlow 101U peristaltic pump and sterile Nalgene™ 50 Platinum-Cured Silicone Tubing (Thermo Scientific™, South Africa). After filtration the polycarbonate filter membranes were transferred and stored in sterile microcentrifuge tubes at -20°C until DNA extraction (Kwon *et al.*, 2011; Pinto *et al.*, 2012), some samples had multiple filters use. Samples were analysed as described in Chapter 3, Section 3.2.4.

4.3 RESULTS AND DISCUSSION

4.3.1 Community membership/composition for Reservoir B

Mean relative abundance (Figure 4.1) shows Proteobacteria to have the highest abundance values across all sample points, regardless of the time of day it was taken. Planctomycetes and Bacteroidetes were the second-highest abundances. These findings are common in the drinking water systems.

4.3.2 Alpha diversity

Figure 4.2 gives an indication of the alpha diversity within the community population. Shannon values are not that high, indicating that there is a lower diversity richness within the community. Observed values show lower values than the previous data set, this could be too low microbial numbers or perhaps sequencing error. Pielou values closer to 100 show a higher evenness, the values, showed a lower evenness. A site with low evenness can indicate that a few species dominate the area. Commonly in a community, if there is a significant disparity between the numbers of the individual species, a low evenness is observed.

Figure 4.3 provides an indication of the alpha diversity within the community population. Shannon values are not that high, indicating a lower diversity richness within the community; however, the inlet source shows a higher richness than the two other points. Observed values show lower values than the previous data set; the microbial numbers could be too low. Again, we see that the source inlet has higher counts. Pielou values closer to 100 show a higher evenness, which showed a lower evenness. A site with low evenness can indicate that a few species dominate the area. Commonly in a community, if there is a significant disparity between the numbers of the individual species, a low evenness is observed; again, the inlet was seen to be more even than the other sample points. This indicates diversity differences between sample points, according to these parameters.

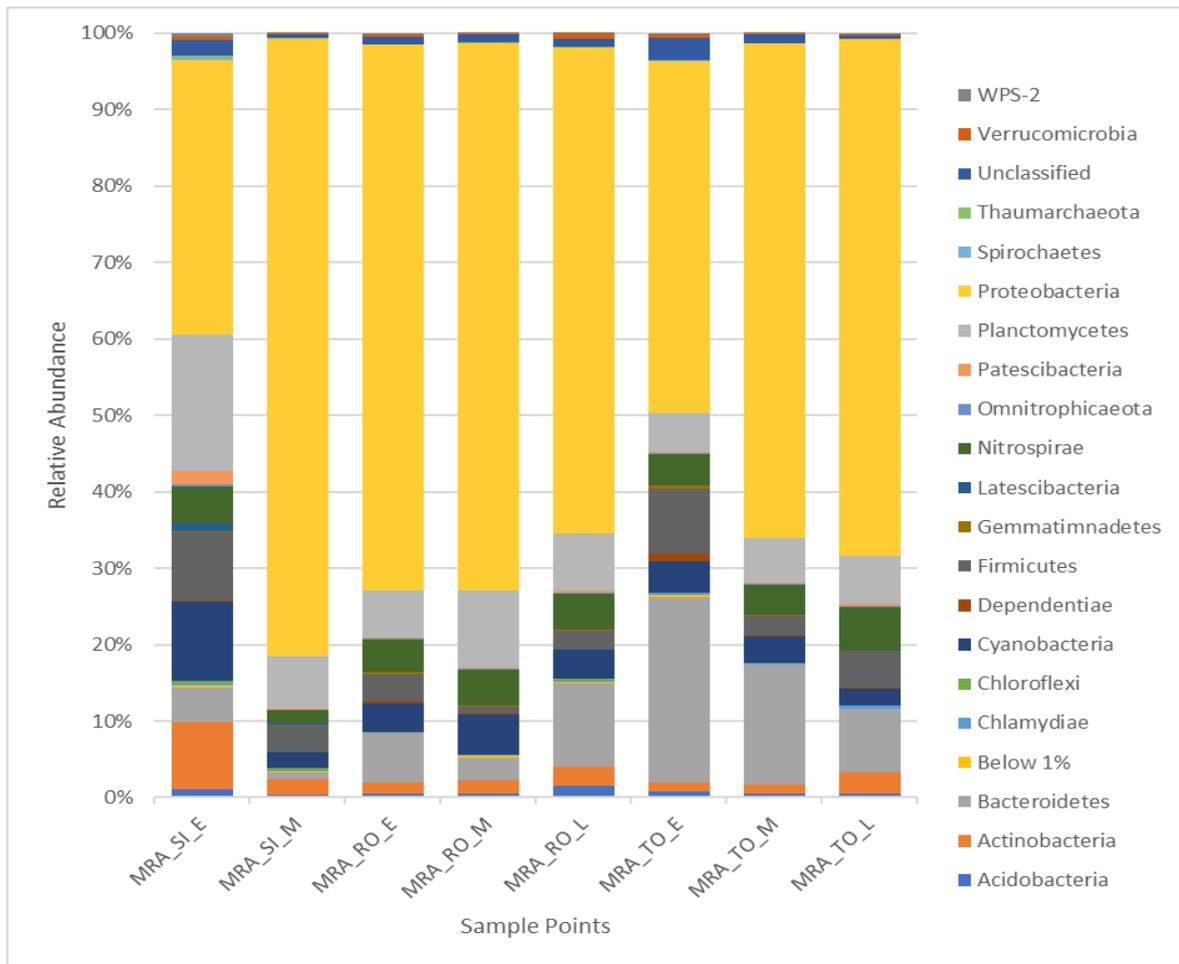


Figure 4.1: Representation of the mean relative abundance (MRA) of bacterial phyla of all sample points. Points include source inlet (SI), reservoir outlet (RO) and tower outlet (TO). Samples were taken early (E), midday(M) and late(L).

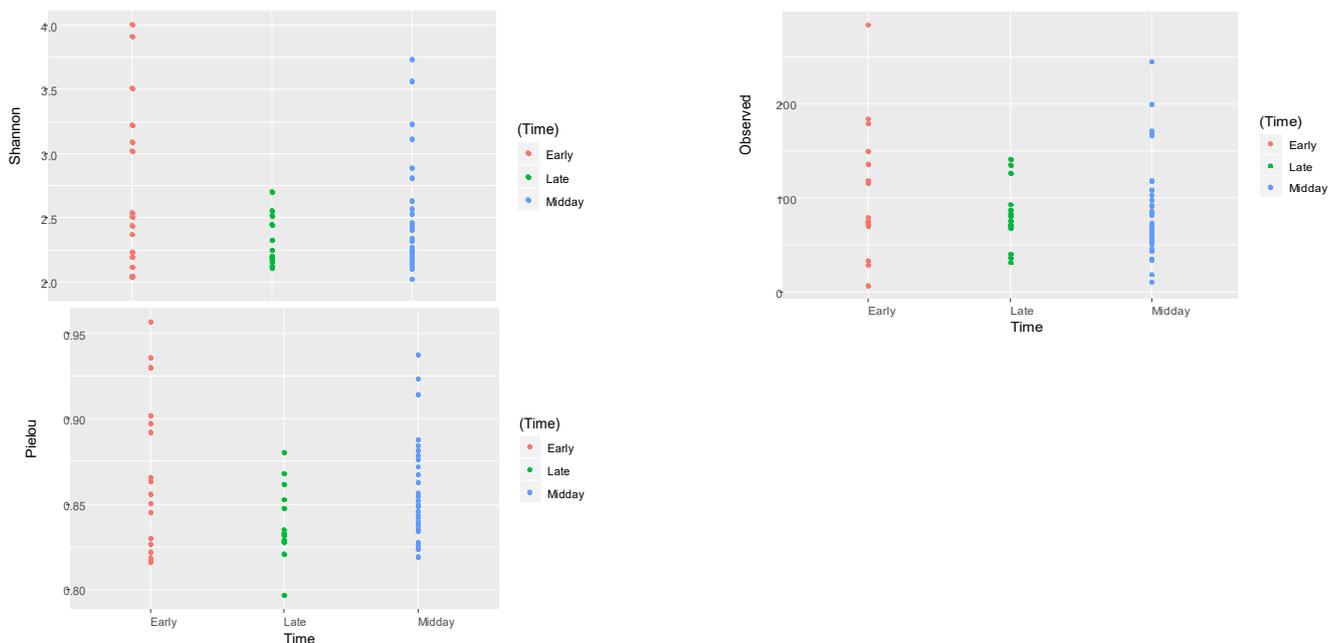


Figure 4.2: Alpha diversity representation of the time of day and alpha parameters. Shannon (top left) observed (top right) and pielou (bottom left).

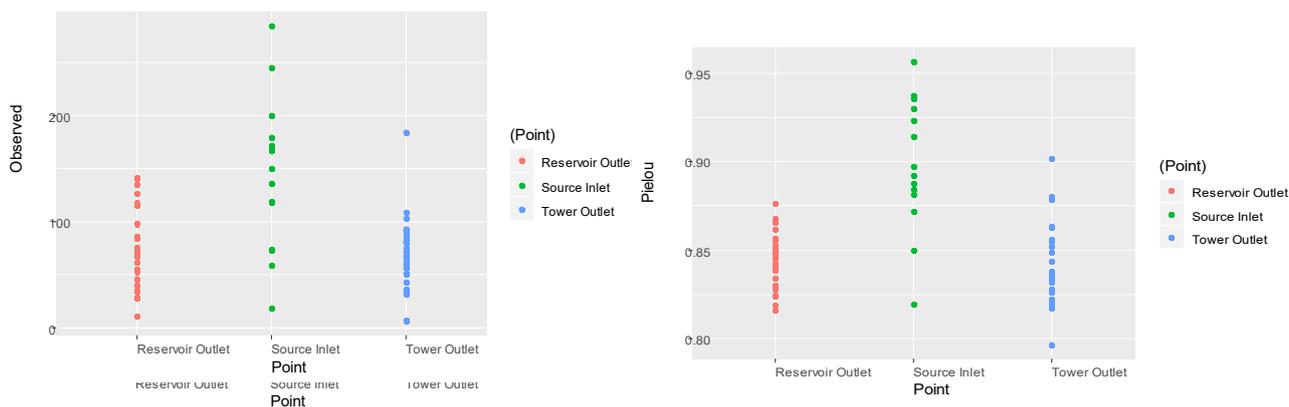


Figure 4.3: Alpha diversity representation of the sample points and alpha diversity parameters. Shannon (top left) observed (top right) and pielou (bottom left).

4.3.3 Beta diversity

The Bray Curtis and Jaccard matrixes in a PcoA plot of the different time points in the sampling data are shown in Figure 4.4 and Figure 4.5. The data shows no distinct groups based on location or time of sampling. The community is quite mixed with little unique diversity within the population.

Figure 4.6 shows a heatmap for the Reservoir B data set representing a constituted >1% of the total sequence abundance amplicon sequence variant table (ASV-SEQ). The redder the blocks the more abundant the ASV. A key for the horizontal axis of the heatmap forms part of Appendix C (Table C2).

The community seemed to be similar regardless of the time of day the samples were taken. This could imply that the retention time (as tested up to 72 hours) didn't influence the community composition. Dominate SEQs seen in all sample points included *Nitrosomonas*, *Phrrobacter*, *Sphingomonas* and *Sphaerotilus*. These bacteria are common in drinking water systems (Simek, 2011). Interestingly *Flavobacterium* was seen to be in high abundance SEQs on days 1-3 and then disappeared on days 4-7. Temperature, nutrient availability and presence of sedimentation are known to influence growth rates of *Flavobacterium* (Madetoja *et al.*, 2003).

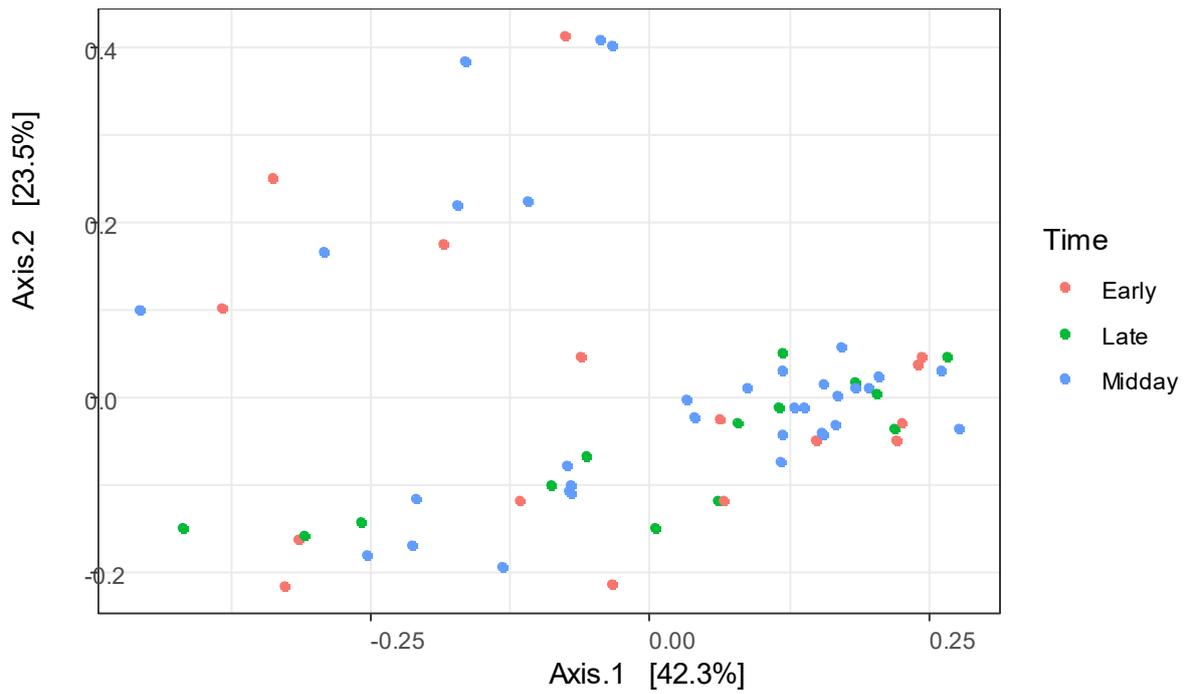


Figure 4.4: PcoA plot, showing Bray Curtis matrix against time of samples taken.

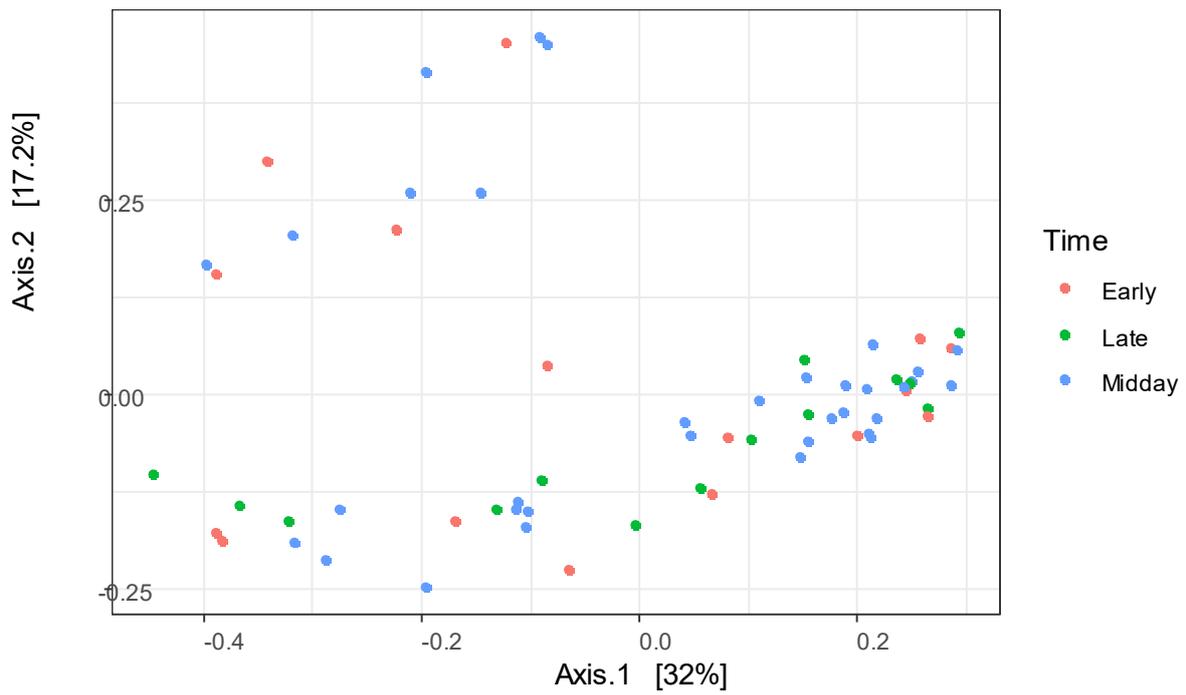


Figure 4.5: PcoA plot, showing Jaccard matrix against time of samples taken.

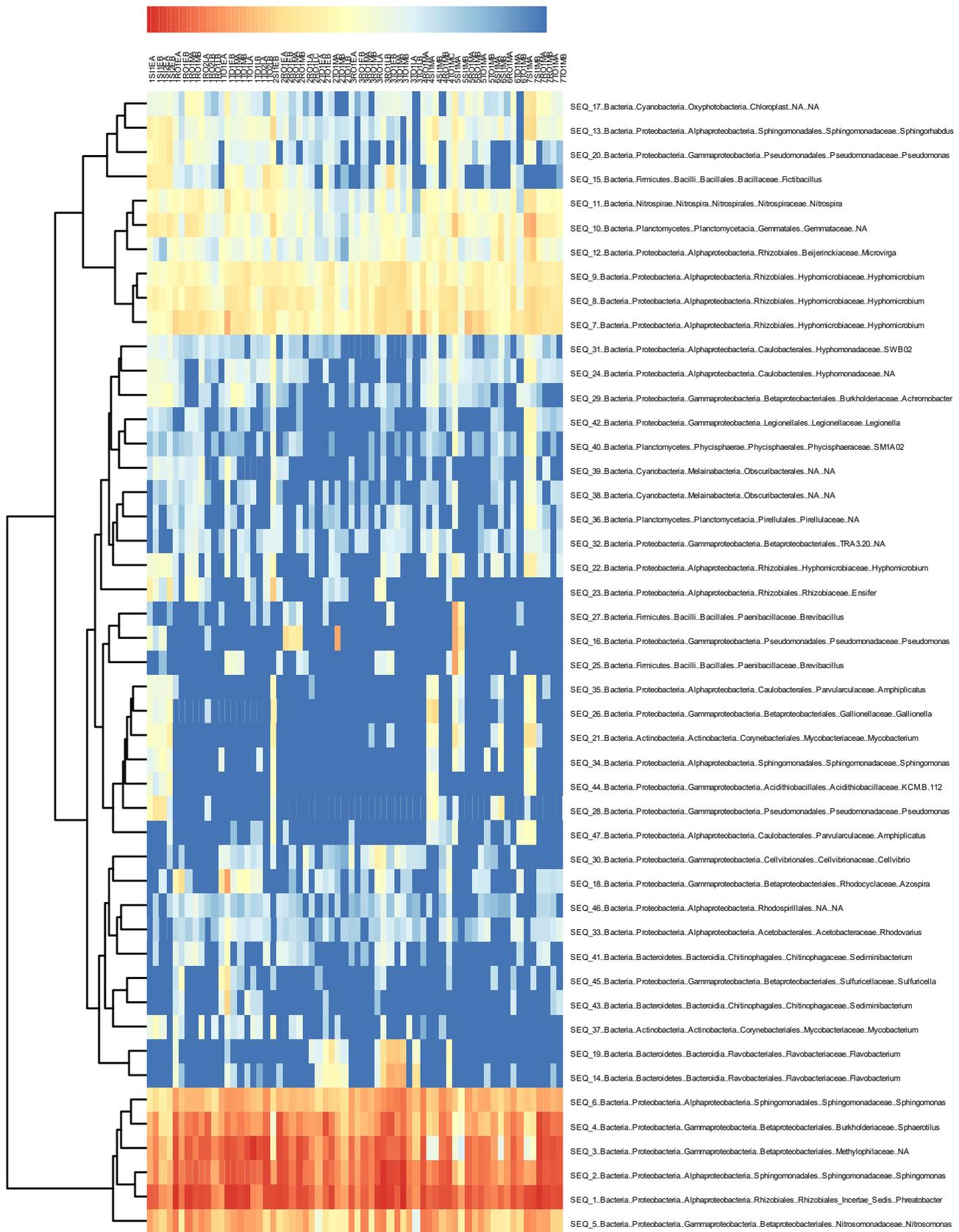


Figure 4.6: Representation of the abundance's levels of different sampling points over the different sampling days (Reservoir B).

CHAPTER 5: CONCLUSIONS

This study provided an overview of the microbial diversity across and within various sampling locations based on 16 S rRNA gene data generated. The study showed that the bacterial diversity is high amongst all these samples and the bacterial community itself is unique among samples and the abundance of these bacteria vary. For the relative abundances, a temporal influence was most likely the reason for the drastic differences in the bacterial diversity across all sampling points. The trends seen across the sampling locations are that the bacterial community differs across each sampling point. When looking at each sampling location, one would expect that the bacterial community present is fairly consistent but a temporal influence as well as external factors such as sampling conditions and distribution system conditions between each sampling point may have played a vital role influencing the latter.

The use of sequencing on this chlorinated drinking water system may however mean that the DNA of numerous dead bacteria were sequenced and this may have completely overshadowed what was truly happening in the system. An alternative would be to adapt the assay to restrict the sequencing to detect only viable organisms, but as there are still issues linked to this approach and it was not attempted. As samples were taken at points along an extensive network it is assumed that most of these organisms could still be viable. The data gathered however confirms the available information in the literature and there are some unique patterns in this dataset present during specific seasons. Overall it can be concluded that there were no specific species that could be solely linked to an increase in bacterial numbers and that each system and sampling point have their own community which changes over time depending on a number of abiotic factors. Further investigation is required to investigate the viable community within the system and the interactions that take place between the microbes. This study however has provided a foundation for local drinking water utilities to know which microorganisms are part of the system and the results confirm the information in the current literature gathered from studies on drinking water utilities around the world.

The isolates obtained in this study were cultured using culture-dependent techniques that most drinking water utilities use presently to monitor the safety and quality of water in the system. The focus of water utilities is to produce operationally safe and aesthetically acceptable water to the consumers that are free from pathogens which pose a health risk. The cultivable bacterial community is very diverse with each location having isolates from a variety of different families. This approach has definite value as it provides undeniable proof of the viability of the bacteria isolated. The same group of species were generally detected with each of the media. A diverse community was present, and several isolates clustered separately from the type strains and formed their own little groups which may indicate the presence of new species. The isolates which grouped together were not all from the same sampling point which further highlighted the diversity amongst the locations. Although the isolates were cultured directly from water samples in this study, the majority were typically found to be associated with other sources based on the primary literature.

Analysis revealed that the community performed all the basic functionality but also showed that some enzymes were not present in any samples; this indicated that regardless of the month sampling was performed, the microbial functionality remained constant throughout, showing microbial stability. It should be noted that enzymes absent in all sample points play a role in biofilm formation, which is good in drinking water as biofilms can harbour microbes. The system studied was cleaned and did not have noticeable biofilm formation inside.

In conclusion, the design and operations of the reservoirs showed to be optimal, as there were only minor changes in the microbial population, showing a stable environment.

REFERENCES

1. Allen, M.J., Edberg, S.C., Reasoner, D.J. 2004. Heterotrophic plate count bacteria – What is their significance in drinking water? *Int. J. Food Microbiol.* 92, 265-274.
2. Asaf, S., Numan, M., Khan, A.L. Al-Harrasi, A. 2020. Sphingomonas: from diversity and genomics to functional role in environmental remediation and plant growth. *Crit. Rev. Biotechnol.* 40, 138-152.
3. Ayling, M., Clark, M.D. and Leggett, R.M. 2019. New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics.* 21,584-594.
4. Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T. and Bartram, J. 2014. Faecal contamination of drinking-water in low and middle-income countries: a systematic review and meta-analysis. *PLoS Medicine.* 5, 644.
5. Bain, R.E., Gundry, S.W., Wright, J.A., Yang, H., Pedley, S. and Bartram, J.K. 2012. Accounting for water quality in monitoring access to safe drinking-water as part of the Millennium Development Goals: lessons from five countries. *Bulletin of the World Health Organization.* 3, 228-235.
6. Basile, N., Fuamba, M. and Barbeau, B. 2008. Optimization of water tank design and location in water distribution systems, in *Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008.* Kruger National Park, South Africa: 361-373..
7. Batte, M., Appenzeller, B.M.R., Grandjean, D., Fass, S., Gauthier, V., Jorand, F., Mathieu, L., Boualam, M., Saby, S., Block, J.C. 2003. Biofilms in drinking water distribution systems. *Rev. Environ. Sci. Bio/Technology* 2, 147-168.
8. Bautista-De los Santos, Q.M., Schroeder, J.L., Sevillano-Rivera, M.C., Sungthong, R., Ijaz, U.Z., Sloan, W.T., Pinto, A.J. 2016. Emerging investigators series: Microbial communities in full-scale drinking water distribution systems-a meta-analysis. *Environ. Sci. Water Res. Technol.* 2, 631-644.
9. Bennett, A.C., Murugapiran, S.K., Hamilton, T.L. 2020. Temperature impacts community structure and function of phototrophic Chloroflexi and Cyanobacteria in two alkaline hot springs in Yellowstone National Park. *Environment Microbiology.* 5, 503-513.
10. Benson, R., Conerly, O.D., Sander, W., Batt, A.L., Boone, J.S., Furlong, E.T. and Simmons, J.E. 2017. Human health screening and public health significance of contaminants of emerging concern detected in public water supplies. *The Science of the Total Environment.* 579,1643-1648.
11. Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.U. and Egli, T. 2007. Assessment and interpretation of bacterial viability by using LIVE/DEAD BacLight kit in combination with flow cytometry. *Applied and Environmental Microbiology.* 73, 3283-3290.
12. Berry, D., Xi, C., Raskin, L. 2006. Microbial ecology of drinking water distribution systems. *Curr. Opin. Biotechnol.* 17, 297-302.
13. Besmer, M.D., Weissbrodt, D.G., Kratochvil, B.E., Sigrist, J.A., Weyland, M.S., Hammes, F. 2014. The feasibility of automated online flow cytometry for in-situ monitoring of microbial dynamics in aquatic ecosystems. *Frontiers in Microbiology.* 5, 265.
14. Beszteri, B., Temperton, B. Frickenhaus, S. and Giovannoni, S.J. 2010. Average genome size: a potential source of bias in comparative metagenomics. *ISME Journal.* 4, 1075-1077.
15. Bifulco, J.M., Shirey, J.J., Bissonnette, G.K. 1989. Detection of *Acinetobacter* spp. in rural drinking water supplies. *Appl. Environ. Microbiol.* 55, 2214-2219.
16. Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30, 2114-2120.
17. Bose, T., Haque, M.M., Reddy, C.V.S.K., Mande, S. and Raghava, G.P.S. 2015. COGNIZER: A framework for functional annotation metagenomic datasets. *PloS One.* 11, 142-145.
18. Bray, J.R. and Curtis, J.T. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs.* 4, 325-349.
19. Brito, J.A., Denkmann, K., Pereira, I.A., Archer, M., Dahl, C. 2015. Thiosulfate dehydrogenase (TsdA) from *Allochromatium vinosum*: structural and functional insights into thiosulfate oxidation. *Journal of*

Biology and Chemistry. 14, 9222-9238.

20. Buse, H.Y., Lu, J., Lu, X., Mou, X., Ashbolt, N.J. 2014. Microbial diversities (16S and 18S rRNA gene pyrosequencing) and environmental pathogens within drinking water biofilms grown on the common premise plumbing materials plasticized polyvinylchloride and copper. *FEMS Microbiol. Ecol.* 88, 280-295.
21. Busse, H.J., Denner, E.B.M., Buczolits, S., Salkinoja-Salonen, M., Bennisar, A., Kämpfer, P. 2003. *Sphingomonas aurantiaca* sp. nov., *Sphingomonas aerolata* sp. nov. and *Sphingomonas faeni* sp. nov., air- and dustborne and Antarctic, orange-pigmented, psychrotolerant bacteria, and emended description of the genus *Sphingomonas*. *Int. J. Syst. Evol. Microbiol.* 53, 1253-1260.
22. Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., A, A.J. 2016. HHS Public Access 13, 581-583.
23. Chandra, S., Saxena, T., Nehra, S., Krishna Mohan, M. 2016. Quality assessment of supplied drinking water in Jaipur city, India, using PCR-based approach. *Environ. Earth Sci.* 75, 1-14.
24. Chao, Y., Mao, Y., Wang, Z., Zhang, T. 2015. Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing. *Sci. Rep.* 5, 1-13.
25. Chaudhary, D.K., Jeong, S.W., Kim, J. 2017. *Sphingobium naphthae* sp. nov., with the ability to degrade aliphatic hydrocarbons, isolated from oil-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 67, 2986-2993.
26. Chaumeil, P.A., Mussig, A.J., Hegenholtz, P. and Parks, D.H. 2019. GTDB-TK: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics.* 36, 1925-1927.
27. Chen, X., Wang, H., Xu, J., Song, D., Sun, G., Xu, M. 2016. *Sphingobium hydrophobicum* sp. nov., a hydrophobic bacterium isolated from electronic-waste-contaminated sediment. *Int. J. Syst. Evol. Microbiol.* 66, 3912-3916.
28. Cheng, Q., Sanglard, D., Vanhanen, S., Liu, H.T., Bombelli, P., Smith, A. and Slabas, A.R. 2005. *Candida* yeast long chain fatty alcohol oxidase is a c-type haemoprotein and plays an important role in long chain fatty acid metabolism. *Molecular and Cell Biology of Lipids.* 1735, 192-203.
29. Choi, G.M., Jo, J.H., Kang, M.S., Kim, M.S., Lee, S.Y., Im, W.T. 2017. *Sphingomonas aquatica* sp. Nov., isolated from tap water. *Int. J. Syst. Evol. Microbiol.* 67, 845-850.
30. Chowdhury, S. 2009. Trihalomethanes in drinking water: Effect of natural organic matter distribution. *Water SA.* 39, 1-6.
31. Codina, R., Principe, J., Munoz, C. and Baiges, J. 2015. Numerical modeling of chlorine concentration in water storage tanks, *International Journal for Numerical Methods in Fluids.* 79, 84-107.
32. Cooksey, R.C., de Waard, J.H., Yakus, M.A., Rivera, I., Chopite, M., Toney, S.R., Morlock, G.P., Butler, W.R. 2004. *Mycobacterium cosmeticum* sp. nov., a novel rapidly growing species isolated from a cosmetic infection and from a nail salon. *Int. J. Syst. Evol. Microbiol.* 54, 2385-2391.
33. Craun, G.F. and Calderon, R.L. 2001. Waterborne disease outbreaks caused by distribution system deficiencies. *Journal of the American Water Works Association.* 93, 64-75.
34. Craun, G.F. and McCabe, Li. 1973. Review of the causes of waterborne-disease outbreaks. *Journal of the American Water Works Association.* 65, 74-84.
35. Delafont, V., Brouke, A., Bouchon, D., Moulin, L., Héchard, Y. 2013. Microbiome of free-living amoebae isolated from drinking water. *Water Res.* 47, 6958-6965.
36. Digiano, F.A. and Zhang, W. 2005. Pipe section reactor to evaluate chlorine-wall reaction. *Journal of the American Water Works Association.* 1, 74-85.
37. Domingo, J.W.S., Ashbolt, N.J., Rice, E.W. 2015. Drinking Water Fourth Edition. Reference Module in Biomedical Sciences. Elsevier.
38. Douterelo, I., Sharpe, R.L., Boxall JB 2013. Influence of hydraulic regimes on bacterial community structure and composition in an experimental drinking water distribution system. *Water Research.* 47, 503-516.
39. Edokpayi, J.N., Rogawski, E.T., Kahler, D.M., Hill, C.L., Reynolds, C., Nyathi, E., Smith, J.A., Odiyo, J.O., Samie, A., Bessong, P. and Dillingham, R. 2018. Challenges to sustainable safe drinking water: A case study of water quality and use across seasons in rural communities in Limpopo province. *Water.*

- 10,159-166.
40. Edwards, J. and Maher, J. 2008. Water quality considerations for distribution system storage facilities. *Journal of the American Water Works Association*. 7, 60-65.
 41. Fard, M.A. and Barkdoll, B.D. 2018. Stagnation reduction in drinking water storage tanks through internal piping with implications for water quality improvement. *Journal of Hydraulic Engineering*. 144, 05018004.
 42. Feinsten, L.M., Sul, W.J., Blackwood, C.B. 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Applied and Environmental Microbiology*. 75, 5428-5433.
 43. Feng, G. Da, Yang, S.Z., Zhu, H.H., Li, H.P. 2018. Emended descriptions of the species *Sphingomonas adhaesiva* Yabuuchi et al. 1990 and *Sphingomonas ginsenosidimutans* Choi et al. 2011. *Int. J. Syst. Evol. Microbiol.* 68, 970-973.
 44. Ferrenberg, S., O'Neill, S. P., Knelman, J. E., Todd, B., Duggan, S., Bradley, D., Robinson, T., Schmidt, S. K., Townsend, A. R., Williams, M. W. 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *The ISME Journal*. 7: 1102-1111.
 45. Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P.J., Soen, Y., Shental, N. 2018. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome*. 6, 1-13.
 46. Gallego, V., García, M.T., Ventosa, A., 2005. *Methylobacterium isbiliense* sp. nov., isolated from the drinking water system of Sevilla, Spain. *Int. J. Syst. Evol. Microbiol.* 55, 2333-2337.
 47. Gensberger, E.T., Gössl, E.-M., Antonielli, L., Sessitsch, A., Kostić, T. 2015. Effect of different heterotrophic plate count methods on the estimation of the composition of the culturable microbial community. *PeerJ* 3, e862.
 48. Gillespie, S., Lipphaus, P., Green, J., Parsons, S., Weir, P., Juskowiak, K., Jefferson, B., Jarvis, P., Nocker, A. 2014. Assessing microbiological water quality in drinking water distribution systems with disinfectant residual using flow cytometry. *Water Research*. 65, 224-234.
 49. Gobet, A., Quince, C., Ramette, A. 2010. Multivariate Cut-off Level Analysis (MultiCoLA) of large community data sets. *Nucleic Acid Research*. 38, e 155.
 50. Godoy, F., Vancanneyt, M., Martínez, M., Steinbüchel, A., Swings, J., Rehm, B.H.A. 2003. *Sphingopyxis chilensis* sp. nov., a chlorophenol-degrading bacterium that accumulates polyhydroxyalkanoate, and transfer of *Sphingomonas alaskensis* to *Sphingopyxis alaskensis* comb. nov. *Int. J. Syst. Evol. Microbiol.* 53, 473-477.
 51. Graham, E.D., Heidelberg, J.F. and Tuly, B.J. 2018. Potential for primary productivity in a globally distributed bacterial phototroph. *Multidisciplinary Journal of Microbial Ecology*. 12, 1861-1866.
 52. Grayman, W.A., Rossman, L.A., Arnold, C., Deininger, R.A., Smith, C., Smith, J.F, and Schniphe, R. 2000. Water quality modelling of distribution storage facilities. American Water Works Association Research Foundation, Denver.
 53. Grayman, W.A., Rossman, L.A., Deininger, R.A., Smith, C.D., Arnold, C.N., and Smith, J.F. 2004. Mixing and aging of water in distribution system storage facilities. *Journal of the American Water Works Association*. 9, 70-80.
 54. Green, P., Bousfield, I. 1983. Emendation of *Methylobacterium* Patt, Cole, and Hanson 1976. *Methylobacterium rhodium* (Heumann 1962) comb. nov. corrig.; *Methylobacterium radiotolerans* (Ito and Iizuka 1971) comb. nov. corrig., *Methylobacterium mesophilicum* (Austin and Goodfellow 1979) comb. nov. *Int. J. Syst. Evol. Microbiol.*, 33, 875-877.
 55. Gualtieri, C. 2009. Analysis of Flow and Concentration Fields in a Baffled Circular Storage Tank, in 33rd IAHR Congress: Water Engineering for a Sustainable Environment.
 56. Hack, D.J. 1984. State regulation of chloramine. *Journal American Water Works Association*. 77, 1-4.
 57. Hallam, N.B., West, J.R., Forest, C.F., Powell, J.C. and Spencer, I. 2002. The decay of chlorine associated with the pipe wall in water distribution systems. *Water Research*. 14, 3479-3488.
 58. Hammes, F., Berney, M., Wang, Y., Vital, M., Koster, O., Egli, T. 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Research*. 44(17), 4868-4877.

59. Hoefel, D., Grooby, W.L., Monis, P.T., Andrews, S., Saint, C.P. 2003. Enumeration of water-borne bacteria using viability assays and flow cytometry: a comparison to culture-based techniques. *Journal of microbiological methods*. 55(3), 585-597.
60. Holinger, E.P., Ross, K.A., Robertson, C.E., Stevens, M.J., Harris, J.K., Pace, N.R. 2014. Molecular analysis of point-of-use municipal drinking water microbiology. *Water Res.* 49, 225-235.
61. Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W., Liu, W.T. 2012. Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl. Environ. Microbiol.* 78, 7856-7865.
62. Jaccard, J. and Turrisi, R. 2003. Interaction effects in multiple regression. Second edition. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-072. Thousand Oaks, CA: Sage.
63. Jia, L., Jiang, B., Huang, F., Hu, X. 2019. Nitrogen removal mechanism and microbial community changes of bioaugmentation subsurface wastewater infiltration system. *Bioresour. Technol.* 294, 122140.
64. Jost, L. 2007. Partitioning diversity into independent alpha and beta components. *Ecology*. 88, 2427-2439
65. Jovel, J., Patterson, J., Wang, W., Hotte, N., Keefe, S.O., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., Wong, G.K. 2016. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7, 1-17.
66. Kaestli, M., O'Donnell, M., Rose, A., Webb, J.R., Mayo, M., Currie, B.J., Gibb, K. 2019. Opportunistic pathogens and large microbial diversity detected in source-to-distribution drinking water of three remote communities in Northern Australia. *PLoS Negl. Trop. Dis.* 13, 1-23.
67. Kanehisa, M., Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27-30.
68. Kang, D., Li, F., Kirton, E.S., Thomas, A., Egan, R.S. and Wang, Z. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *Peer Journal*. 10, 275-285.
69. Kelly, J.J., Minalt, N., Culotti, A., Pryor, M., Packman, A. 2014. Temporal variations in the abundance and composition of biofilm communities colonizing drinking water distribution pipes. *PLoS One*. 9, e098542.
70. Kennedy, R.H. 2001. Considerations for establishing nutrient criteria for reservoirs, lakes, and reservoir management. *Dams and Reservoirs*. 3, 175-187.
71. Khan, S., Melville, B.W., Shamseldin, A.Y. and Fischer, C. 2012. Investigation of flow patterns in storm water retention ponds using CFD. *Journal of Environmental Engineering*. 139, 61-69.
72. Kirmeyer, G.J., Kirby, L., Murphy, B.M., Noran, F., Martel, K.D., Lund, T.W., Anderson, J.L. and Medhurst, R. 1999. Maintaining and Operating finished water storage facilities. Denver, Colorado.
73. Kirmeyer, G.J., Odell, L.H., Jacangelo, J.J., Wilczac, A. and Wolfe, R. 1995. Nitrification occurrence and control in chloraminated water systems. American Water Works Association Research Foundation, Denver.
74. Korajkic, A., Parfrey, L.W., McMinn, B.R., Baeza, Y.V., VanTeuren, W., Knight, R., Shanks, O.C. 2015. Changes in bacterial and eukaryotic communities during sewage decomposition in Mississippi river water. *Water Res.* 69, 30-39.
75. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the Miseq. *Applied and Environmental Microbiology*. 79, 5112-5120.
76. 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*. 2, 149-155.
77. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA Sequences to the human genome. *Genome Biology*. 10, 25.

78. Lautenschlager, K., Hwang, C., Liu, W.-T., Boon, N., Köster, O., Vrouwenvelder, H., Egli, T., Hammes, F. 2013. A microbiology-based multiparametric approach towards assessing biological stability in drinking water distribution networks. *Water Research*. 47, 3015-3025.
79. LeChevallier, M.W., Welch, N.J. and Smith, D.B. 1996. Full-scale studies of factors related to coliform regrowth in drinking water. *Applied Environmental Microbiology*. 62, 2201-2211.
80. Leclerc, H., Schwartzbrod, L., Dei-Cas, E. 2002. Microbial agents associated with waterborne diseases. *Crit. Rev. Microbiol.* 28, 371-409.
81. Li, Weiyang, Wang, F., Zhang, J., Qiao, Y., Xu, C., Liu, Y., Qian, L., Li, Wenming, Dong, B. 2016. Community shift of biofilms developed in a full-scale drinking water distribution system switching from different water sources. *Sci. Total Environ.* 544, 499-506.
82. Li, Y., Gao, Y., Zhang, W., Wang, C., Wang, P., Niu, L. and Wu, H. 2019. Homogeneous selection dominates the microbial community assembly in the sediment of the three gorges reservoir. *Science of The Total Environment*. 690, 50-60.
83. Ling, F., Hwang, C., LeChevallier, M.W., Andersen, G.L., Liu, W.T. 2016. Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system. *ISME J.* 10, 582-595.
84. Ling, F., Whitaker, R., LeChevallier, M.W., Liu, W.T. 2018. Drinking water microbiome assembly induced by water stagnation. *The ISME Journal*, 12, 1520.
85. Liu, G., Verberk, J.Q.J.C. and Van Dijk, J.C. 2013. Bacteriology of drinking water distribution systems: An integral and multidimensional review. *Applied Microbiology Biotechnology*. 97, 9265-9276.
86. Lliros, 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.
87. Lu, J., Buse, H.Y., Gomez-Alvarez, V., Struewing, I., Santo Domingo, J., Ashbolt, N.J. 2014. Impact of drinking water conditions and copper materials on downstream biofilm microbial communities and *Legionella pneumophila* colonization. *J. Appl. Microbiol.* 117, 905-918.
88. Lu, P., Chen, C., Wang, Q., Wang, Z., Zhang, X., Xie, S. 2013. Phylogenetic diversity of microbial communities in real drinking water distribution systems. *Biotechnol. Bioprocess Eng.* 18, 119-124.
89. Lührig, K., 2016. Bacterial communities in drinking water biofilms. Doctoral dissertation, University of Lund.
90. Lui, G., Verbeck, J. Q. J. C, van Dijk, J. C. 2013. Bacteriology of drinking water distribution systems: an integral and multidimensional review. *Applied and Environmental Microbiology*. 97, 9265-9276.
91. Luyt ,C., Tandlich, R., Muller, W. J. and Wilhelmi, B.S. 2012. Microbial monitoring of surface water in South Africa: an overview. *International Journal of Environmental Research and Public Health*. 9, 2669-2693.
92. Madetoja, J., Nystedt, S. and Wiklund, T. 2003. Survival and virulence of *Flavobacterium psychrophilum* in water microcosms. *Microbiology Ecology*. 43, 217-223.
93. Mahmood, F. 2005. Evaluation of water mixing characteristics in distribution system storage Tanks. *Journal of the American Water Works Association*. 97, 3-74.
94. Mara, D.D., Horan, N.J., 2003. *The handbook of water and wastewater microbiology*. Elsevier.
95. Marek, M., Stoesser, T., Roberts, P.J.W., Weitbrecht, V. and Jirka, G.H. 2007. CFD modelling of turbulent jet mixing in a water storage tank. *Proceedings of the Congress-International Association for Hydraulic Research*. 32, 551.
96. Martel, K.D., Kirmeyer, G.J., Murphy, B.M., Paul, F., Kirby, L., Lund, T.W., Anderson, J.L., Medhurst, R., Murphy, B.M. and Noran, P.F. 2002. Preventing Water Quality Deterioration in Finished Water Storage Facilities. *Journal – American Water Works Association*. 94, 139-148.
97. Martínez-Solano, F.J., Iglesias Rey, P.L., Gualtieri, C. and López-Jiménez, P.A. 2010. Modelling flow and concentration field in rectangular water tanks. In *International Congress on Environmental Modelling and Software Modelling for Environment's Sake, iEMSs, Fifth Biennial Meeting*. Ottawa, Canada.

98. McNaughton, K.J. and Sinclair, C.G. 1966. Submerged Jets in Short cylindrical flow vessels. *Journal of Fluid Mechanics*. 2, 367-375.
99. Messner, M., Shaw, S., Regli S., Rotert, K. and Soller, J. 2006. An approach for developing a national estimate of waterborne diseases due to drinking water and a national estimate model application. *Journal of Water and Health*. 4, 201-240.
100. Moncho-Esteve, I.J., Palau-Salvador, G., Brevis, W., Vaas, M.O. and López-Jiménez, P.A. 2015. Numerical simulation of the hydrodynamics and turbulent mixing process in a drinking water storage tank. *Journal of Hydraulic Research*. 53, 207-217.
101. Nescerecka, A., Rubulis, J., Vital, M., Juhna, T. and Hammes, F. 2015. Biological instability in a chlorination drinking water distribution network. *PLoS One*. 5, 3-4.
102. Nordblom, O. and Bergdahl, L. 2003. Initiation of Stagnation in Drinking Water Storage Tanks. *Journal of Hydraulic Engineering*. 130, 49-57.
103. Pal, R., Bhasin, V.K., Lal, R. 2006. Proposal to reclassify *Sphingomonas xenophaga* Stolz et al. 2000 and *Sphingomonas taejonensis* Lee et al. 2001 as *Sphingobium xenophagum* comb. nov. and *Sphingopyxis taejonensis* comb. nov., respectively. *Int. J. Syst. Evol. Microbiol.* 56, 667-670.
104. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P-A., Hugenholtz, P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. 36, 996-1004.
105. Patt, T.E., Cole, G.C., Hanson, R.S. 1976. *Methylobacterium*, a New Genus of Facultatively Methylophilic Bacteria. *Int. J. Syst. Bacteriol.* 26, 226-229.
106. Pinto, A. J., Xi, C and Raskin, L. 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environmental Science and Technology*. 46: 8851-8859.
107. Potgieter, S., Pinto, A., Sigudu, M., Ncube, E., Venter, S. 2018. Long-term spatial and temporal microbial community dynamics in a large-scale drinking water distribution system with multiple disinfectant regimes. *Water Res.* 139, 406-419.
108. Powell, J.C, Hallam, N.B., West, J.R., Forster, C.F. and Simms, J. 2000. Factors which control bulk chlorine decay rates. *Procedia Engineering*. 1, 117-126.
109. Prest, E.I., El-Chakhtoura, J., Hammes, F., Saikaly, P.E., van Loosdrecht, M.C.M., Vrouwenvelder, J.S. 2014. Combining flow cytometry and 16S rRNA gene pyrosequencing: A promising approach for drinking water monitoring and characterization. *Water Res.* 63, 179-189.
110. Prest, E.I., Hammes, F., van Loosdrecht, M.C.M. and Vrouwenvelder, J.S. 2016a. Biological stability of drinking water: controlling factors, methods, and challenges. *Frontiers in Microbiology*. 7, 45
111. Prest, E.I., Weissbrodt, D.G., Hammes, F., Loosdrecht, M.C. and Vrouwenvelder, J.S. 2016b. Long-Term Bacterial Dynamics in a Full-Scale Drinking Water Distribution System *PLoS ONE*. 11, e0164445.
112. Props, R., Rubbens, P., Besmer, M., Buyschaert, B., Sigrüst, J., Weilenmann, H., Waegeman, W., Boon, N., Hammes, F. 2018. Detection of microbial disturbances in a drinking water microbial community through continuous acquisition and advanced analysis of flow cytometry data. *Water Res.* 145, 73-82.
113. Reisch, C.R., Moran, M.A. and Whitman, W.B. 2011. Bacterial catabolism dimethylsulfoniopropionate (DSMP). *Frontiers in Microbiology*. 2,172.
114. Revetta, R.P., Pemberton, A., Lamendella, R., Iker, B., Santo Domingo, J.W. 2010. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water Res.* 44, 1353-1360.
115. Rittmann, B.E., Snoeyink, V.L. 1984. Achieving biologically stable drinking water. *Journal-American Water Works Association*. 76(10), 106-114.
116. Roberts, P.J.W., Tian, X., Fotis, S. and Duer M. 2006. Physical modeling of mixing in water storage tanks. Denver, C.O.: AWWA Research Foundation.
117. Rossman, L.A. and Grayman, W.M. 1999. Scale model studies of mixing in drinking water storage tanks. *Journal of Environmental Engineering*. 8, 755-761.
118. Rossman, L.A., Uber, J.G. and Grayman, W.M. 1995. Modeling disinfectant residuals in drinking-water.

- Journal of Environmental Engineering. 121, 752-755.
119. Rozej, A., Cydzik-Kwiatkowska, A., Kowalska, B., Kowalski, D. 2015. Structure and microbial diversity of biofilms on different pipe materials of a model drinking water distribution systems. *World J. Microbiol. Biotechnol.* 31, 37-47.
 120. Rubuli 2010. MUWS (Microbiology in Urban Water Systems) – an interdisciplinary approach to study microbial communities in urban water systems. *Drink. Water Eng. Sci.* 3, 91-99.
 121. SABS. 2015. South African National Standard 241-1: Drinking water, Part 1: Microbiological, physical aesthetic and chemical determinants. South African Bureau of Standards, Pretoria.
 122. Sawyer, A.H. Shi, F., Kirby, J.T. and Michael, H.A. 2003. Dynamic response of surface water-groundwater exchange to currents, tides, and waves in shallow estuary. *Advancing Earth and Space Science.* 4, 1749-1758.
 123. Scott, B.A., Pepper I.L., 2010. Water distribution systems as living ecosystems: Impact on taste and odor. *J. Environ. Sci. Heal. – Part A Toxic/Hazardous Subst. Environ. Eng.* 45, 890-900.
 124. Shar, A.H., Kazi, Y.F., Kanhar, N.A., Soomro, I.H. 2012. Bacterial community patterns of municipal water of Sukkur city in different seasons. *African J. Biotechnol.* 11, 2287-2295.
 125. SLMB. 2012. Determining the total cell count and ratios of high and low nucleic acid content cells in freshwater using flow cytometry. Analysis method 333.1, Swiss Food Book. Federal Office of Public Health, Swiss Confederation.
 126. Takeuchi, M., Hamana, K., Hiraishi, A. 2001. Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.* 51, 1405-1417.
 127. Tallon, P.A.M., Magajna, B., Lofranco, C., Leung, K.A.M.T.I.N. 2005. Microbial Indicators of Faecal Contamination in Water: A Current Perspective Ensuring the safety of drinking water is an ongoing process. In developed countries, drinking water regulations require the monitoring of numerous chemical and microbiologic. *Water, Air Soil Pollut.* 166, 139-166.
 128. Tamames, J. and Puente-Sanchez 2019. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology.* 9, 3349.
 129. Tamas, I., Smirnova, A. V., He, Z., Dunfield, P.F. 2014. The (d)evolution of methanotrophy in the Beijerinckiaceae — a comparative genomics analysis. *ISME J.* 8, 369-382.
 130. Tian, R., Ning, D. and He, Z. 2020. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome.* 8, 51
 131. Tian, X. and Roberts, P.J.W. 2008. Mixing in water storage tanks. I: No buoyancy effects. *Journal of Environmental Engineering.* 134, 974-985.
 132. Tortoli, E., Kohl, T.A., Brown-Elliott, B.A., Trovato, A., Leão, S.C., Garcia, M.J., Vasireddy, S., Turenne, C.Y., Griffith, D.E., Phillely, J. V., Baldan, R., Campana, S., Cariani, L., Colombo, C., Taccetti, G., Teri, A., Niemann, S., Wallace, R.J., Cirillo, D.M. 2016. Emended description of *Mycobacterium abscessus*, *Mycobacterium abscessus* subsp. *abscessus* and *Mycobacterium abscessus* subsp. *bolletii* and designation of *Mycobacterium abscessus* subsp. *massiliense* comb. nov. *Int. J. Syst. Evol. Microbiol.* 66, 4471-4479.
 133. Toth, E.M., Vengring, A., Homonnay, Z.G., Keki, Z., Sproere, C., Borsodi, A.K., Marialigeti, K., Schumann, P. 2014. *Phreatobacter oligotrophus* gen. nov., sp. nov., an alphaproteobacterium isolated from ultrapure water of the water purification system of a power plant. *Int. J. System. Evol. Microbiol.* 64, 839-845.
 134. Tsao, H.F., Scheikl, U., Herbold, C., Indra, A., Walochnik, J., Horn, M. 2019. The cooling tower water microbiota: Seasonal dynamics and co-occurrence of bacterial and protist phylotypes. *Water Res.* 159, 464-479.
 135. Turton, A. 2008. Three strategic water quality challenges that decision-makers need to know about and how the CSIR should respond. Science real and relevant: 2nd CSIR Biennial Conference, CSIR International Convention Centre Pretoria, 17-18 November 2008.
 136. Turton, A. 2008. Three strategic water quality challenges that parliamentarians need to know about.

CSIR Report No. CSIR/NRE/WR/IR/2008/0079/C prepared for the October Parliamentary Briefing. Pretoria. Council for Scientific and Industrial Research.

137. Ulrich, N., Kirchner, V., Drucker, R., Wright, J.R., McLimans, C.J., Hazen, T.C., Campa, M.F., Grant, C.J., Lamendella, R. 2018. Response of Aquatic Bacterial Communities to Hydraulic Fracturing in Northwestern Pennsylvania: A Five-Year Study. *Sci. Rep.* 8, 1-13.
138. Urakawa, H., Martens-Habbena, W. and Stahl, DA. 2010. High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Applied and Environmental Microbiology.* 76, 2129-2135.
139. Van Nevel, S., Koetzsch, S., Proctor, C.R., Besmer, M.D., Prest, E.I., Vrouwenvelder, J.S., Knezev, A., Boon, N., Hammes, F. 2017. Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring. *Water Res.* 113, 191-206.
140. Van Zyl, J.E. and Haarhoff, J. Reliability analysis of municipal storage reservoirs using stochastic analysis. 2007. *Journal of the South African Institution of Civil Engineering.* 49, 3-9.
141. Vaz-Moreira, I., Egas, C., Nunes, O.C., Manaia, C.M. 2013. Bacterial diversity from the source to the tap: A comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. *FEMS Microbiol. Ecol.* 83, 361-374.
142. Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology.* 85: 183-206. 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.
143. Wang, H., Hu, C., Li, X. 2015. Characterization of biofilm bacterial communities and cast-iron corrosion in bench-scale reactors with chloraminated drinking water. *Eng. Fail. Anal.* 57, 423-433.
144. 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.
145. Williams, M.M., Domingo, J.W.S., Meckes, M.C., Kelty, C.A., Rochon, H.S. 2004. Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *J. Appl. Microbiol.* 96, 954-964.
146. Wingender, J., Flemming, H.C. 2011. Biofilms in drinking water and their role as reservoir for pathogens. *Int. J. Hyg. Environ. Health* 214, 417-423.
147. World Health Organization 2004. Microbial fact sheets. In: *Guidelines for drinking water quality. Volume 1: Health criteria and other supporting information.* World Health Organization.
148. World Health Organization 2005. *Nutrients in drinking water.* (No. WHO/SDE/WSH/05.09). World Health Organization.
149. World Health Organization 2011. *Guidelines for drinking-water quality, Fourth edition.* World Health Organization.
150. Wu, H., Zhang, J., Mi, Z., Xie, S., Chen, C., Zhang, X. 2015. Biofilm bacterial communities in urban drinking water distribution systems transporting waters with different purification strategies. *Appl. Microbiol. Biotechnol.* 99, 1947-1955.
151. Wu, Y.W., Simmons, B.A. and Singer, S.W. 2015. Maxbin 2.0: An Automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics.* 32, 605-607.
152. Yang, Y., Li, S., Gao, Y., Chen, Y., Zhan, A. 2019. Environment-driven geographical distribution of bacterial communities and identification of indicator taxa in Songhua River. *Ecol. Indic.* 101, 62-70.
153. Young, C.C., Ho, M.J., Arun, A.B., Chen, W.M., Lai, W.A., Shen, F.T., Rekha, P.D., Yassin, A.F. 2007. *Sphingobium olei* sp. nov., isolated from oil-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 57, 2613-2617.
154. Zeng, D.-N., Fan, Z.-Y., Chi, L., Wang, X., Qu, W.-D., Quan, Z.-X. 2013. Analysis of the bacterial communities associated with different drinking water treatment processes. *World J. Microbiol. Biotechnol.* 29, 1573-1584.
155. Zhang, J.M., Khoo, B.C., Teo, C.P., Haja, N., Tham, T.K., Zhong, L. and Lee, H.P. 2013. Passive and active methods for enhancing water quality of service reservoir. *Journal of Hydraulic Engineering.* 139,

745-753.

156. Zhang, W., DiGiano, F.A. 2002. Comparison of bacterial regrowth in distribution systems using free chlorine and chloramine: A statistical study of causative factors. *Water Research*. 36, 1469-1482.
157. Zhang, Y. 2012. *Microbial Community Dynamics and Assembly: Drinking Water Treatment and Distribution*. PhD dissertation, University of Tennessee – Knoxville.
158. Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., Alvarez-Cohen, L. 2015. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio*. 2, 5-6.
159. Zhu, Y., Wang, H., Li, X., Hu, C., Yang, M., Qu, J. 2014. Characterization of biofilm and corrosion of cast iron pipes in drinking water distribution system with UV/Cl₂ disinfection. *Water Research*. 60, 174-181.

APPENDIX A: SUPPLEMENTARY DATA FOR CHAPTER 2

Table A1: Percentage of the taxonomic classification on a family level of the 87 OTUs across 89 sampling points represented by the Heatmap in Figure 2.38

Family	% Abundance
Moraxellaceae	21,1
Burkholderiaceae	18,0
Sphingomonadaceae	14,6
Rhizobiales_Incertae_Sedis	8,2
Pseudomonadaceae	8,1
Unclassified	7,3
Nitrosomonadaceae	5,0
Nitrospiraceae	2,8
Methylophilaceae	2,2
Beijerinckiaceae	1,9
Hyphomicrobiaceae	1,8
Gemmataceae	1,1
Rhodobacteraceae	1,1
Acetobacteraceae	1,0
Chitinophagaceae	0,9
Rhodocyclaceae	0,9
Xanthobacteraceae	0,8
TRA3-20	0,8
Alteromonadaceae	0,4
Spirosomaceae	0,4
Hyphomonadaceae	0,4
Mycobacteriaceae	0,3
Omnitrophaceae	0,3
Mitochondria	0,2
Parvularculaceae	0,2
Holophagaceae	0,1
Caulobacteraceae	0,1
Blastocatellaceae	0,1

Table A2: Percentage of the taxonomic classification on a genus level of the 87 OTUs across 89 sampling points represented by the Heatmap in Figure 2.38

Genus	% abundance
Acinetobacter	20,48
Unclassified	13,43
Sphingomonas	10,68
Phreatobacter	8,17
Pseudomonas	8,09
Delftia	5,53
Nitrosomonas	4,96
Acidovorax	4,94
Nitrospira	2,83
Ralstonia	2,44
Hyphomicrobium	1,82
Sphingorhabdus	1,59
Hydrogenophaga	1,26
Microvirga	1,26
Rhodobacter	0,81
Sphingopyxis	0,80
Porphyrobacter	0,78
Massilia	0,76
Sediminibacterium	0,75
Curvibacter	0,71
Methylobacterium	0,64
Enhydrobacter	0,59
Bradyrhizobium	0,50
Noviherbaspirillum	0,49
Piscinibacter	0,46
DSSF69	0,43
Rhodovarius	0,38
Rheinheimera	0,38
Arcicella	0,38
Limnobacter	0,35
Roseomonas	0,33
Mycobacterium	0,29
Undibacterium	0,27
Candidatus_Omnitrophus	0,26
Cereibacter	0,25
Zoogloea	0,23
Methyloversatilis	0,21
Roseateles	0,21
Lacibacter	0,20
Polaromonas	0,17
Novosphingobium	0,17
Amphiplicatus	0,16
Fimbrigliobus	0,15
Geothrix	0,15
Phenylobacterium	0,14
JGI_0001001-H03	0,13

Table A3: The number and percentages of the 87 OTUs across 89 sampling points represented by the Heatmap in Figure 2.40

OTU	Number of occurrences	% Occurrence
OTU1	45	50,6
OTU2	81	91,0
OTU3	69	77,5
OTU4	38	42,7
OTU5	59	66,3
OTU6	75	84,3
OTU7	33	37,1
OTU8	56	62,9
OTU9	21	23,6
OTU10	30	33,7
OTU11	29	32,6
OTU12	45	50,6
OTU13	25	28,1
OTU14	29	32,6
OTU15	15	16,9
OTU16	30	33,7
OTU17	49	55,1
OTU18	13	14,6
OTU19	49	55,1
OTU20	48	53,9
OTU21	16	18,0
OTU22	28	31,5
OTU23	14	15,7
OTU24	44	49,4
OTU25	17	19,1
OTU26	8	9,0
OTU27	20	22,5
OTU28	16	18,0
OTU29	9	10,1
OTU30	41	46,1
OTU31	21	23,6
OTU32	5	5,6
OTU33	6	6,7
OTU34	3	3,4
OTU35	15	16,9
OTU36	20	22,5
OTU37	21	23,6
OTU38	16	18,0
OTU40	35	39,3
OTU41	10	11,2
OTU42	19	21,3
OTU43	26	29,2
OTU44	8	9,0

OTU	Number of occurrences	% Occurrence
OTU45	15	16,9
OTU46	18	20,2
OTU47	13	14,6
OTU48	24	27,0
OTU49	19	21,3
OTU50	26	29,2
OTU51	16	18,0
OTU52	8	9,0
OTU54	1	1,1
OTU55	24	27,0
OTU56	7	7,9
OTU57	12	13,5
OTU58	3	3,4
OTU59	6	6,7
OTU63	6	6,7
OTU64	24	27,0
OTU66	23	25,8
OTU67	16	18,0
OTU68	2	2,2
OTU69	4	4,5
OTU70	17	19,1
OTU71	15	16,9
OTU73	18	20,2
OTU75	15	16,9
OTU78	3	3,4
OTU79	16	18,0
OTU81	11	12,4
OTU85	14	15,7
OTU86	10	11,2
OTU87	1	1,1
OTU90	5	5,6
OTU95	8	9,0
OTU96	5	5,6
OTU98	9	10,1
OTU104	8	9,0
OTU107	4	4,5
OTU108	10	11,2
OTU109	10	11,2
OTU110	6	6,7
OTU111	4	4,5
OTU117	5	5,6
OTU120	7	7,9
OTU121	9	10,1
OTU127	2	2,2

Table A4: Taxonomic classification of the 87 most abundant OTUs across all sampling locations represented by the Heatmap in Figure 2.40

OTU	Phylum	Class	Order	Family	Genus	Species
OTU1	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Unclassified
OTU2	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Rhizobiales Incertae Sedis	Phreatobacter	Unclassified
OTU3	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified
OTU4	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Delftia	Unclassified
OTU5	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	Unclassified
OTU6	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Nitrosomonadaceae	Nitrosomonas	oligotropha
OTU7	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Acidovorax	Unclassified
OTU8	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified
OTU9	Planctomycetes	Planctomycetacia	Planctomycetales	Unclassified	Unclassified	Unclassified
OTU10	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Ralstonia	Unclassified
OTU11	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Methylophilaceae	Unclassified	Unclassified
OTU12	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira	lenta
OTU13	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	Unclassified
OTU14	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Unclassified
OTU15	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	Unclassified
OTU16	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Hydrogenophaga	Unclassified
OTU17	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Beijerinckiaceae	Microvirga	Unclassified
OTU18	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Unclassified
OTU19	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Hyphomicrobiaceae	Hyphomicrobium	Unclassified
OTU20	Planctomycetes	Planctomycetacia	Gemmatales	Gemmataceae	Unclassified	Unclassified
OTU21	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Unclassified
OTU22	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Porphyrobacter	Unclassified
OTU23	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	Sediminibacterium	goheungense
OTU24	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	TRA3-20	Unclassified	Unclassified
OTU25	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	Unclassified
OTU26	Cyanobacteria	Melainabacteria	Obscuribacterales	Unclassified	Unclassified	Unclassified
OTU27	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Curvibacter	Unclassified
OTU28	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
OTU29	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Beijerinckiaceae	Methylobacterium	Unclassified
OTU30	Planctomycetes	Planctomycetacia	Planctomycetales	Unclassified	Unclassified	Unclassified

OTU31	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Moraxellaceae	Enhydrobacter	aerosaccus
OTU32	Cyanobacteria	Melainabacteria	Obscuribacterales	Unclassified	Unclassified	Unclassified
OTU33	Proteobacteria	Gamma-Proteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	Unclassified
OTU34	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Massilia	Unclassified
OTU35	Planctomycetes	Planctomycetacia	Planctomycetales	Unclassified	Unclassified	Unclassified
OTU36	Proteobacteria	Alpha-Proteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacter	Unclassified
OTU37	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Piscinibacter	aquaticus
OTU38	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis	Unclassified
OTU40	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Xanthobacteraceae	Bradyrhizobium	Unclassified
OTU41	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Noviherbaspirillum	Unclassified
OTU42	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	hankyongensis
OTU43	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Hyphomicrobiaceae	Hyphomicrobium	Unclassified
OTU44	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Rhodocyclaceae	Unclassified	Unclassified
OTU45	Chloroflexi	Anaerolineae	Unclassified	Unclassified	Unclassified	Unclassified
OTU46	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira	Unclassified
OTU47	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	DSSF69	Unclassified
OTU48	Proteobacteria	Alpha-Proteobacteria	Acetobacterales	Acetobacteraceae	Rhodovarius	Unclassified
OTU49	Proteobacteria	Alpha-Proteobacteria	Acetobacterales	Acetobacteraceae	Roseomonas	Unclassified
OTU50	Proteobacteria	Gamma-Proteobacteria	Alteromonadales	Alteromonadaceae	Rheinheimera	Unclassified
OTU51	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Unclassified	Unclassified
OTU52	Bacteroidetes	Bacteroidia	Cytophagales	Spirosomaceae	Arcicella	rosea
OTU54	Proteobacteria	Alpha-Proteobacteria	Caulobacterales	Hyphomonadaceae	Unclassified	Unclassified
OTU55	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Hyphomicrobiaceae	Hyphomicrobium	Unclassified
OTU56	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Limnobacter	thiooxidans
OTU57	Proteobacteria	Alpha-Proteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacter	Unclassified
OTU58	Proteobacteria	Alpha-Proteobacteria	Acetobacterales	Acetobacteraceae	Unclassified	Unclassified
OTU59	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira	Unclassified
OTU63	Proteobacteria	Alpha-Proteobacteria	Rhizobiales	Xanthobacteraceae	Unclassified	Unclassified
OTU64	Acidobacteria	Subgroup_6	Unclassified	Unclassified	Unclassified	Unclassified
OTU66	Actinobacteria	Actinobacteria	Corynebacterales	Mycobacteriaceae	Mycobacterium	Unclassified
OTU67	Cyanobacteria	Oxyphotobacteria	Chloroplast	Unclassified	Unclassified	Unclassified
OTU68	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis	witflariensis
OTU69	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Acidovorax	Unclassified

OTU70	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Undibacterium	macrobrachii
OTU71	Omnitrophicaeota	Omnitrophia	Omnitrophales	Omnitrophaceae	Candidatus_Omnitrophus	Unclassified
OTU73	Proteobacteria	Alpha-Proteobacteria	Rhodobacterales	Rhodobacteraceae	Cereibacter	Unclassified
OTU75	Cyanobacteria	Melainabacteria	Obscuribacterales	Unclassified	Unclassified	Unclassified
OTU78	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Rhodocyclaceae	Zoogloea	caeni
OTU79	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	Lacibacter	Unclassified
OTU81	Proteobacteria	Alpha-Proteobacteria	Rickettsiales	Mitochondria	Unclassified	Unclassified
OTU85	Cyanobacteria	Melainabacteria	Obscuribacterales	Unclassified	Unclassified	Unclassified
OTU86	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Rhodocyclaceae	Methyloversatilis	universalis
OTU87	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Roseateles	Unclassified
OTU90	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Massilia	Unclassified
OTU95	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Burkholderiaceae	Polaromonas	Unclassified
OTU96	Planctomycetes	Planctomycetacia	Planctomycetales	Unclassified	Unclassified	Unclassified
OTU98	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium	subterraneum
OTU104	Proteobacteria	Alpha-Proteobacteria	Caulobacterales	Parvularculaceae	Amphiplicatus	Unclassified
OTU107	Planctomycetes	Planctomycetacia	Gemmatales	Gemmataceae	Fimbrioglobus	Unclassified
OTU108	Planctomycetes	Planctomycetacia	Planctomycetales	Unclassified	Unclassified	Unclassified
OTU109	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	Nitrosomonadaceae	Nitrosomonas	Unclassified
OTU110	Acidobacteria	Holophagae	Holophagales	Holophagaceae	Geothrix	fermentans
OTU111	Proteobacteria	Alpha-Proteobacteria	Sphingomonadales	Sphingomonadaceae	Unclassified	Unclassified
OTU117	Proteobacteria	Alpha-Proteobacteria	Caulobacterales	Caulobacteraceae	Phenylobacterium	koreense
OTU120	Proteobacteria	Gamma-Proteobacteria	Betaproteobacteriales	TRA3-20	Unclassified	Unclassified
OTU121	Acidobacteria	Blastocatellia (Subgroup 4)	Blastocatellales	Blastocatellaceae	JGI_0001001-H03	Unclassified
OTU127	Cyanobacteria	Melainabacteria	Obscuribacterales	Unclassified	Unclassified	Unclassified

APPENDIX B: SUPPLEMENTARY DATA FOR CHAPTER 3

Table B1: Key to the labels used for individual samples in Figure 3.31

Group	Label on heatmap	Group	Label on heatmap	Group	Label on heatmap
Mar_1_1m	1	Aug_3_1m	29	Sep_R_5_3m	57
Mar_R_1_1m	2	Aug_R_3_1m	30	Sep_R_Inlet	58
Mar_1_3m	3	Aug_3_3m	31	Sep_R_Outlet	59
Mar_R_1_3m	4	Aug_R_3_3m	32	Oct_1	60
Mar_2_1m	5	Aug_4_1m	33	Oct_2	61
Mar_2_3m	6	Aug_R_4_1m	34	Oct_inlet	62
Mar_3_1m	7	Aug_R_4_3m	35	Nov_1_1m	63
Mar_R_3_1m	8	Aug_R_5_1m	36	Nov_1_3m	64
Mar_3_3m	9	Aug_R_5_3m	37	Nov_2_1m	65
Mar_R_3_3m	10	Aug_R_Inlet	38	Nov_2_3m	66
Mar_4_1m	11	Aug_Outlet	39	Nov_3_1m	67
Mar_R_4_1m	12	Aug_R_Outlet	40	Nov_3_3m	68
Mar_4_3m	13	Sep_1_1m	41	Nov_4_1m	69
Mar_R_4_3m	14	Sep_R_1_1m	42	Nov_4_3m	70
Mar_5_1m	15	Sep_1_3m	43	Nov_5_1m	71
Mar_R_5_1m	16	Sep_R_1_3m	44	Nov_5_3m	72
Mar_5_3m	17	Sep_2_1m	45	Nov_Inlet	73
Mar_R_5_3m	18	Sep_R_2_1m	46	Nov_R_Inlet	74
Mar_Inlet	19	Sep_2_3m	47	Nov_Outlet	75
Mar_R_Inlet	20	Sep_R_2_3m	48	Nov_R_Outlet	76
Mar_Outlet	21	Sep_3_1m	49	Dec_1_1m	77
Mar_R_Outlet	22	Sep_R_3_1m	50	Dec_2_1m	78
Aug_1_1m	23	Sep_3_3m	51	Dec_3_1m	79
Aug_R_1_1m	24	Sep_R_3_3m	52	Dec_4_1m	80
Aug_1_3m	25	Sep_4_1m	53	Dec_5_1m	81
Aug_R_1_3m	26	Sep_4_3m	54	Dec_Inlet	82
Aug_2_1m	27	Sep_5_1m	55		
Aug_2_3m	28	Sep_R_5_1m	56		

Table B2: Key to the sample code used for individual samples in Figure 3.37

Sample	Day	Point	Duplicate	Time	Filter
1SI1EA	1	Source Inlet	1	Early	A
1SI1EB	1	Source Inlet	1	Early	B
1SI2EA	1	Source Inlet	2	Early	A
1SI2EB	1	Source Inlet	2	Early	B
1RO1EA	1	Reservoir Outlet	1	Early	A
1RO1EB	1	Reservoir Outlet	1	Early	B
1RO1MA	1	Reservoir Outlet	1	Midday	A
1RO1MB	1	Reservoir Outlet	1	Midday	B
1RO2LA	1	Reservoir Outlet	2	Late	A
1RO2LB	1	Reservoir Outlet	2	Late	B
1RO1LB	1	Reservoir Outlet	1	Late	B
1TO1EA	1	Tower Outlet	1	Early	A
1TO1EB	1	Tower Outlet	1	Early	B
1TO1MA	1	Tower Outlet	1	Midday	A
1TO1MB	1	Tower Outlet	1	Midday	B
1TO1LA	1	Tower Outlet	1	Late	A
1TO1LB	1	Tower Outlet	1	Late	B
1TO2LA	1	Tower Outlet	2	Late	A
1TO2LB	1	Tower Outlet	2	Late	B
2RO1EA	2	Reservoir Outlet	1	Early	A
2RO1EB	2	Reservoir Outlet	1	Early	B
2RO1MA	2	Reservoir Outlet	1	Midday	A
2RO1MB	2	Reservoir Outlet	1	Midday	B
2RO1LA	2	Reservoir Outlet	1	Late	A
2RO1LC	2	Reservoir Outlet	1	Late	C
2TO1EA	2	Tower Outlet	1	Early	A
2TO1EB	2	Tower Outlet	1	Early	B
2TO1MA	2	Tower Outlet	1	Midday	A
2TO1MB	2	Tower Outlet	1	Midday	B
2TO1LB	2	Tower Outlet	1	Late	B
3RO1EA	3	Reservoir Outlet	1	Early	A

Table B2 (Continue): Key to the sample code used for individual samples in Figure 3.37

Sample	Day	Point	Duplicate	Time	Filter
3RO1EB	3	Reservoir Outlet	1	Early	B
3RO1MA	3	Reservoir Outlet	1	Midday	A
3RO1MB	3	Reservoir Outlet	1	Midday	B
3RO1LA	3	Reservoir Outlet	1	Late	A
3RO1LB	3	Reservoir Outlet	1	Late	B
3TO1EB	3	Tower Outlet	1	Early	B
3TO1MA	3	Tower Outlet	1	Midday	A
3TO1MB	3	Tower Outlet	1	Midday	B
3TO1LA	3	Tower Outlet	1	Late	A
3TO1LB	3	Tower Outlet	1	Late	B
4RO1MA	4	Reservoir Outlet	1	Midday	A
4SI1MA	4	Source Inlet	1	Midday	A
4SI1MB	4	Source Inlet	1	Midday	B
4RO1MB	4	Reservoir Outlet	1	Midday	B
4TO1MC	4	Tower Outlet	1	Midday	C
5SI1MA	5	Source Inlet	1	Midday	A
5SI1MB	5	Source Inlet	1	Midday	B
5RO1MA	5	Reservoir Outlet	1	Midday	A
5RO1MB	5	Reservoir Outlet	1	Midday	B
5TO1MA	5	Tower Outlet	1	Midday	A
5TO1MB	5	Tower Outlet	1	Midday	B
6SI1MA	6	Source Inlet	1	Midday	A
6SI1MB	6	Source Inlet	1	Midday	B
6RO1MA	6	Reservoir Outlet	1	Midday	A
6TO1MA	6	Tower Outlet	1	Midday	A
6TO1MB	6	Tower Outlet	1	Midday	B
7SI1MA	7	Source Inlet	1	Midday	A
7SI1MB	7	Source Inlet	1	Midday	B
7RO1MA	7	Reservoir Outlet	1	Midday	A
7RO1MB	7	Reservoir Outlet	1	Midday	B
7TO1MA	7	Tower Outlet	1	Midday	A
7TO1MB	7	Tower Outlet	1	Midday	B