Pyrosequencing analysis of roof-harvested rainwater and river water used for domestic purposes in Luthengele village in the Eastern Cape Province of South Africa

Lizyben Chidamba[1a], Lise Korsten[1b]

[1]Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa

[a]Corresponding author: Cell: 0782540386; Email: lizybenc@gmail.com

[b]Cell: +27 79 522 8476; Phone: +27 12 420 3295; Email: lise.korsten@up.ac.za

Abstract

Pyrosequencing targeting the V1-V3 hypervariable of the 16S rDNA was used to investigate the bacterial diversity in river and roof-harvested rainwater (RHRW) used for potable purposes by rural households in Luthengele village in the Eastern Cape Province of South Africa. The phylum *Proteobacteria* dominated the data set (80.5% of all reads), while 4.2% of the reads could not be classified to any of the known phyla at a probability of 0.8 or higher (unclassified bacteria). At class level, the classes; *Betaproteobacteria* (50.4% of all reads), *Alphaproteobacteria* (16.2%), *Verrucomicrobiae* (6.6%), *Planctomycetacia* (5.7%) and *Sphingobacteria* (3%) dominated the data set in all the samples. Although the class Verrucomicrobiae constituted 6.6% of all sequences, 88.6% of the sequences were from the river sample where the class represented 43.7% of the observed sequences in the sample. The bacteria community structure clearly showed significant similarities between RHRW and differences with the river water control sample, suggesting different levels of contamination and environmental factors affecting the various water sources. Moreover, signatures of potential pathogens including *Legionella, Acinetobacter, Pseudomonas, Clostridia, Chromobacterium, Yersinia* and *Serratia* were detected, and the proportions of *Legionella* were relatively higher suggesting a potential health risk to households using RHRW. This work provides guidance for prioritizing subsequent culturable and quantitative analysis, to ensure that potentially significant pathogens are not left out of risk estimations.

Keywords: rainwater harvesting, contamination, faecal indicator, pathogens, pyrosequencing

Introduction

In most South African rural areas and informal settlements municipal water is either inconsistently supplied or is not available at all. Hence, households rely on roof-harvested rainwater (RHRW) for domestic purposes including drinking, food preparation, bathing and washing (Kahinda et al. 2010). Although, there is a general public health perception that RHRW is safe to drink, the presence of potential pathogens such as *E. coli, Aeromonas* spp., *Campylobacter* spp., *Salmonella* spp., *Legionella pneumophila, Giardia* spp., *Cryptosporidium* spp., *Vibrio* spp. and enteric viruses, have been reported in this water source as reviewed by Ahmed et al. (2011).

When rain falls it is initially free of microorganisms but acquires them on its way to the earth's surface. Microbes can be present in dust particles and in the air which then become incorporated into the rainwater (Kaushik et al. 2012; 2014). At the catchment surface of a rainwater harvesting system (RWH), dust, debris and faecal droppings that have been deposited by animals or wind are washed into the tank together with the microorganisms they contain (Ahmed et al. 2012a). Inside the storage tank, microorganisms may die with time or survive and proliferate in biofilms (Rasid 2009). Of the microorganisms that can be present in RHRW, bacteria are abundant and diverse and may include both pathogenic and non-pathogenic species whose presence and prevalence are indicative of the factors contributing to the observed microbial quality (Ahmed et al. 2012b). While efforts are ongoing to minimise contamination in RHRW, successful mitigation is limited unless the variety and concentrations of contaminating pathogens are known. It is therefore important to develop tools that can effectively detect and characterise pathogens of concern in RHRW (Shanks et al. 2010).

Most microbiological characterisation methods for drinking water are specified by local, national or international norms and depend on conventional culturing methods, including heterotrophic plate counts and selective culturing for specific faecal bacteria groups or species (Ahmed et al. 2012b). Among the pathogens transmitted in water, enteric pathogens have been the focus of RHRW quality management (Sinclair et al. 2008). Surveys of pathogens in water have thus far only included a limited number of known infectious agents and indicators (Ahmed et al. 2011). This however, may underestimate the total pathogen

content and diversity and, thus limit the extent to which one can fully understand potential infectious risks from RHRW use.

Reports on the microbial quality of harvested rainwater (HRW) have generally not considered the total microbial population (Ahmed et al. 2011). However, phylogenetic identification of microorganisms within given environments and correlation of their observed diversity and distribution is critical in understanding HRW quality (Kaushik et al. 2014). In particular sources of microbial pollution in the form of faecal material, dust and biofilm formation and microbial proliferation in the holding tank are important factors to consider to better understand microbial populations in RHRW (Ahmed et al. 2010a).

In order to characterise and describe risks posed by the use of RHRW, a comprehensive understanding of the pathogen diversity and abundance is required. However, water microbial communities cannot be adequately studied by traditional culture-based techniques. Consequently, a variety of molecular fingerprinting methods have been developed and include denaturing gradient gel electrophoresis (DGGE), 16S rRNA clone libraries and fluorescence *in situ* hybridisation (FISH) (Luxmy et al. 2000; Meays et al. 2004; Nocker et al. 2007). Of these methods, the construction of 16S rRNA clone libraries appears to give the highest sequencing depth. Although the resolution is typically limited to less than 10^3 sequence identifications per sample (DeSantis et al. 2007). This sampling depth is not good enough to identify less common pathogens, of which are typically present at concentrations far less than 1% proportion of the environmental microbial populations (Bibby et al. 2010). In contrast, the recently developed high-throughput next-generation sequencing, such as 454 pyrosequencing of the small sub-unit (SSU) rRNA genes offers an alternative, in which detailed community structure can be achieved in combination with a fairly high taxonomic resolution (Margulies et al. 2005).

Not only does high-throughput next-generation sequencing allow for the detection and identification of dominant bacteria phylotype profiles within a sample but the high sequence numbers produced allows for the detection of rare species including pathogens within bacterial communities (Sogin et al. 2006). Currently, the technique has been used to explore

human microbiomes (Costello et al. 2009) and environmental samples such as soil (Roesch et al. 2007) and oceans (Sogin et al. 2006). A few investigations have been performed on environmental water samples from either waste or drinking water (Cottrell et al. 2005; Martiny et al. 2005; Poitelon et al. 2009). The aim of this study was to assess microbial communities in RHRW using samples from a village with a typical rural setting that includes rearing of a variety of domestic animals. Furthermore, the applicability of the pyrosequencing analysis technique to detect potential pathogens in RHRW was evaluated.

Methodology

Study site and sample collection

A total of nine water samples were collected in 2L plastic containers in duplicates directly from seven randomly selected RHRW tanks and two individual samples from river water that is also used for potable purposes by the villagers in Luthengele, Port St Johns, Eastern Cape Province, South Africa and water that had been collected from a RHRW tank (tank 2) and stored in the kitchen prior to use (point of use (POU)) (herein after referred to as kitchen water or T-2-POU). The two individual samples were included for comparative purposes. This was intended to provide scoping data on the microbial variations that may exist between river water and RHRW both used by the villagers for potable purposes. We further intended to provide scoping data on the influence of water handling practice on RHRW microbiological quality. The study site, Luthengele village in the Eastern Cape Province of South Africa is situated in a mountainous area with steep terrains. The area has no boreholes and the households depend on RHRW during the wet season, which is supplemented with stream water in dryer months. Generally water is fetched from RHRW tanks and stored in smaller containers in the kitchen prior to use.

DNA extraction and pyrosequencing

Bacteria in water samples (1L) were concentrated onto 0.45µm cellulose nitrate filters (Sartorious Stedim Biotech, Goettingen, Germany) by vacuum filtration (Shi, et al. 2013). Filters were stored at -20^{o}C for less than two weeks prior to DNA extraction and pyrosequencing analysis. DNA was extracted from filters using the Quick-gDNA Miniprep

kit (ZymoReseach, USA) using the manufacturer's protocol. DNA samples were sent for tag encorded pyrosequencing at Inqaba Labs (Pretoria, South Africa) using the using a Roche GS FLX+ 454 pyrosequencer (GATC Biotech, Konstanz, Germany). The 16S rRNA gene was amplified using universal bacterial primer set 27F (5'-AGAGTTTGATCMTGGCTCAG-3') (Weisburg et al. 1991) and 518R (5'-ATTACCGCGGCTGCTGG-3') (Muyzer et al. 1993).

Pyrosequencing data processing and analysis

Sequences that were of low quality or shorter than 150 bp in length were removed from the pyrosequencing-derived data sets using the UCLUST pyrosequencing pipeline initial processing (Cleary et al. 2012). The web-server SnoWMAn 1.15 (http://snowman.genome.tugraz.at) (Stocker et al. 2010) was used with the following settings for the taxonomic-based analysis: analysis type: UCLUST pipeline; chimera filtering reference database: Chimerarefdbs_09_Feb_2012_aligned; Mothur alignment reference database: fungalITS_24-Sep-2013-current_prokMSA_aligned; Mothur chimera checking reference database: chimerarefdbs_39-Feb-2012_16S_aligned and taxonomic reference database: greengenes_24-Mar-2010; rarefaction method: RDP; taxonomy: RDP; confidence threshold: 80%; include taxa covering more than 1%. Operational taxonomic units (OTUs) for rarefaction analysis, were clustered at 3% (species level), 5% (genus level) and 20% (phylum level) dissimilarity cut-offs (Hur and Chun 2004). The calculation of ACE, Evenness, Shannon, Chao1 and Species Richness diversity indices and the generation of principal coordinate analysis (PCoA) plots were generated using the open source software package QIIME (http://qiime.sourceforge.net), designed for the analysis of high-throughput community sequencing data (Caporaso et al. 2010).

Pathogen identification

To identify pathogens, sorted and trimmed sequences that had been classified to belong to pathogenic genera at 80% confidence threshold were selected. These genus classifications were examined for the presence of sequences of pathogenic species known to be present in water (Rusin et al. 1997). To achieve this, sequences were identified against their closest relatives in the NCBI Genbank 16s rRNA database. Closely related sequences were downloaded and aligned. The 16S rRNA gene sequences of representative pathogenic species

extracted from the NCBI Genbank were aligned together with our sequences from pathogenic genera and clustered in MEGA5.2 (Tamura et al. 2011). The cluster dendrogram was generated with the following settings in MEGA5.2: Neighbour-Joining statistical method, Bootstrap method (1000 replications) as a test of phylogeny and Jukes-Cantor model (Jukes and Cantor 1969). The distance matrix output from MEGA5.2 constructed with Jukes-Cantor model was examined for sample sequences that were within 0.03 Jukes-Cantor distance of a known pathogen sequence. Sequences within the limit distance relative to a pathogen were classified as potential pathogens (Bibby et al. 2010)

Results

Sequencing depth

Pyrosequencing of the nine samples resulted in 18510 reads that passed the Roche 454 quality control algorithm. Of these, 14618 reads passed trimming and filtering of the SnoWMAn UCLUST pipeline. A total of 14057 sequences passed uniquing step and were used for alignment and chimera filtering, 3101 (22.1%) sequences were identified as chimeras and were excluded from downstream analysis, leaving 10956 sequences. On average, an individual sample harboured 1265 reads (range 855-1739 reads).

The 10956 identified unique sequences were clustered at 94% similarity into 1092 OTUs (Keijser et al., 2008). Of the 1092 observed OTUs, 95.6% could be classified into their respective phylum. However, the number of sequences that could be classified gradually declined from 84.7% at class level, 69.3% order level to 41% at family level. At genus level, a number of higher classification groups had completely disappeared and only 22.2% of OTUs could be classified. To overcome the challenge of decreasing numbers of classifiable sequences as one moves from phylum to species level, we compared our samples at phylum, order, class, family and genus.

Alpha diversity

To determine richness and diversity, OTUs were identified at genetic distances of 3% (species level), 5% (genus level) and 20% (phylum level) by using quality sequences with a

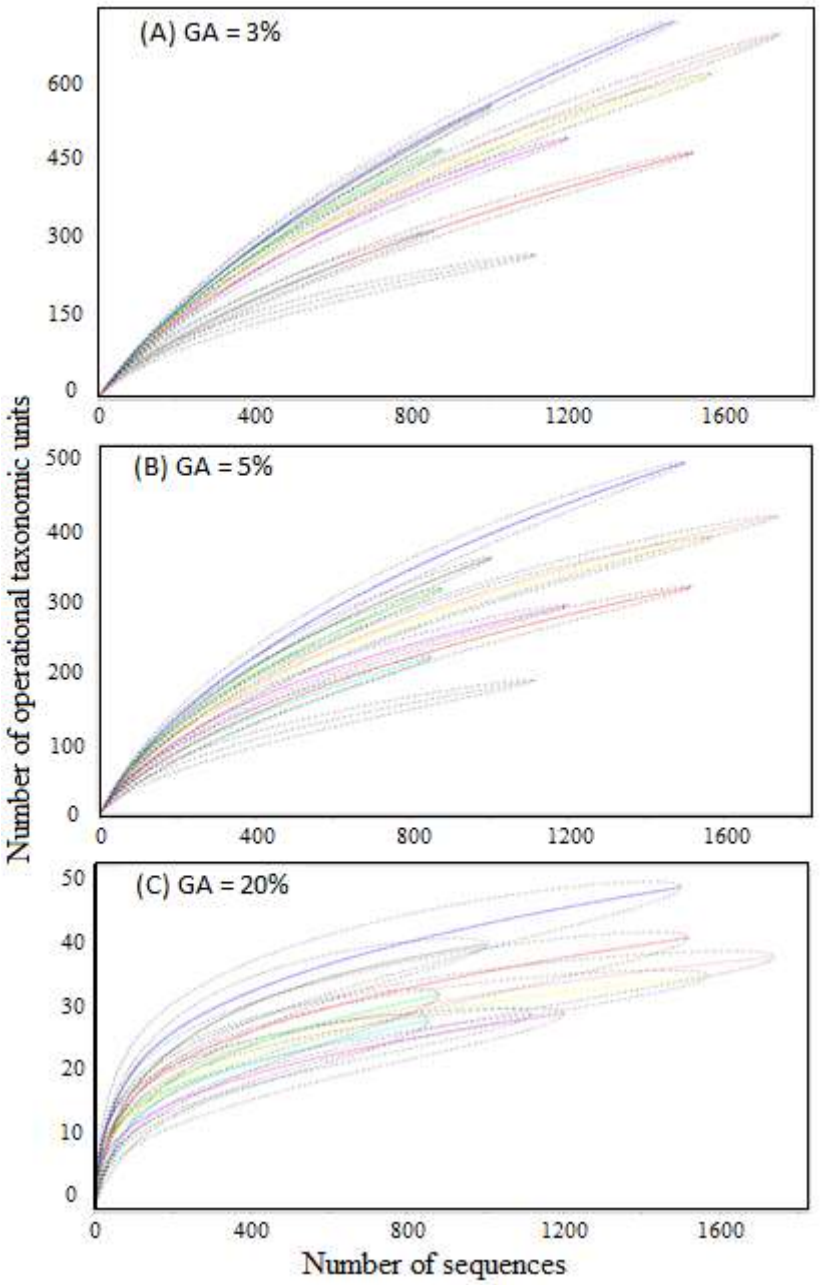

Figure 1: Rarefication curves for species richness and diversity at genetic distances of 3% (A) (species level), 5% (B) (genus level) and 20% (C) (phylum level).

read length of at least 150bp per sample (Pankhurst et al. 2012; Dowd et al. 2008). At 20% sequence divergence, rarefaction curves for all samples showed saturation, indicating that the surveying effort covered almost the full extent of phylum level of taxonomic diversity (Figure 1). However, the rarefaction curves of observed OTUs at genus and species level did not plateau with increasing reads suggesting that a higher number of reads per sample would have provided more comprehensive data of the bacteria taxa. Although the rarefaction curves

did not fully plateau, they had begun to plateau at ~800 reads indicating that we were able to capture the main components of the samples bacterial community diversity at the species and genus level with our level of sequencing depth.

The Chao1, species richness estimator, Shannon and ACE diversity indices indicated that in general, the most diverse bacterial populations were found in water samples from tanks 1, 6 and 7 and the least in the water sample from tank 5 and the kitchen (Table 1). The Chao1 richness varied between 529.3 and 1708.3, the Shannon index between 5.05 and 5.96, and species richness between 276 and 740.

Table 1: Alpha diversity parameters of bacterial communities from river and roof harvested rainwater samples.

Sample ID	ACE	Evenness	Shannon	Chao1	Species Richness
River	850.34	0.83	5.09	850.04	475
T-2-POU	321.62	0.90	5.05	334.02	276
Tank-1	1019.36	0.89	5.88	974.75	740
Tank-2	584.03	0.91	5.60	580.06	482
Tank-3	799.65	0.89	5.76	831.23	629
Tank-4	577.72	0.91	5.68	603.17	503
Tank-5	373.42	0.90	5.17	378.89	322
Tank-6	873.91	0.91	5.96	868.63	705
Tank-7	760.87	0.89	5.66	793.07	566

Phylum level diversity

At phylum level 95.6% of sequences were classified into ten phyla of the domain bacteria (Figure 2). *Proteobacteria* dominated the data set (80.5% of all reads), while 4.2% of the reads could not be classified to any of the known phyla at a probability of 0.8 or higher (unclassified bacteria), 8.7% of reads were classified as *Plancomycetes*, 4.1% as *Bacteroidetes*, 0.73% as *Acidobacteria*, 0.73% as *Verrucomicrobia*, 0.53% as *Armatimonadetes*, 0.46% as *Actinobacteria*, 0.07% as *Nitrospira*, 0.02% as *Firmicutes* and 0.01% as *Gemmatimonadetes*.

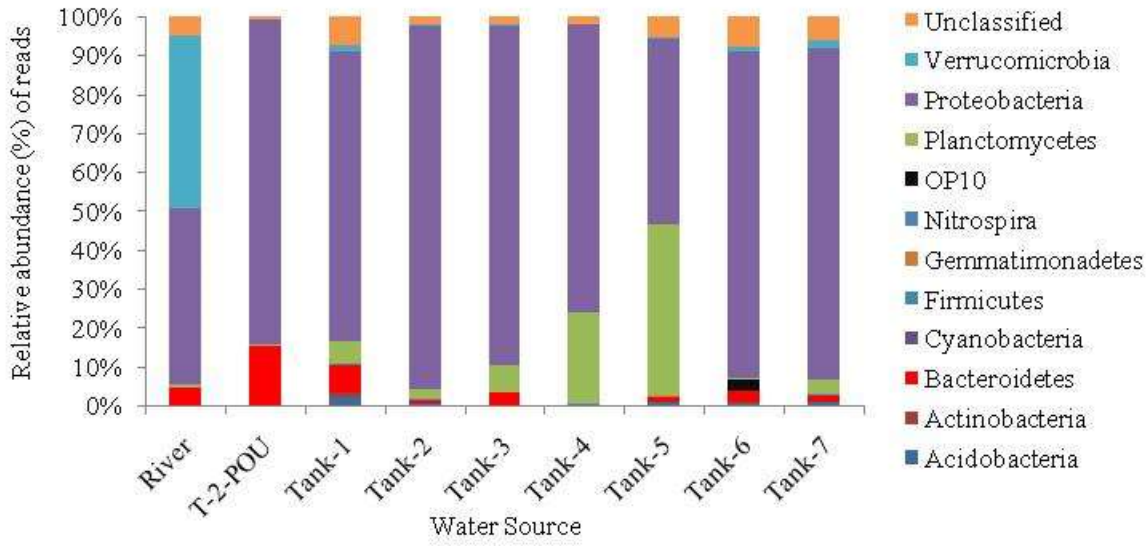

Figure 2: Average relative abundance (%) of the reads classified into the predominant microbial phyla by individual samples.

Proteobacteria were the most dominant phyla in all the samples. The phylum *Verrucomicrobia* (44.5%) was similarly dominant as *Proteobacteria* (45.3%) in river water, although it was detected at less than 1% in other samples. Phylum *Armatimonadetes* was only detected in tank 6. *Planctomycetes* were dominant in tank water samples and tank 4 (23.2%) and 5 (44%) had the highest abundance. Other phyla with interesting distribution included *Gemmatimonadetes* detected only in tank 5 (0.12%), *Firmicutes* in tank 1 (0.13%), *Cyanobacteria* in river water (0.72%) and *Nitrospira* in tank 2 (0.34%), tank 4 (0.08%) and tank 7 (0.1%). *Bacteroidetes* were mainly detected in T-2-POU water (15.6%) but were not detected in tank 2, where the water was collected from. *Bacteroidetes* were also detected in stream (4.4%), tank 1 (7.4%), tank 3 (3.6%), tank 5 (1.2%), 6 (2.7%) and 7 (1.8%).

Class level diversity

At class level, 85.6% of sequences were classified into sixteen classes (Figure 3). The classes; *Betaproteobacteria* (50.4% of all reads), *Alphaproteobacteria* (16.2%) *Verrucomicrobiae* (6.6%), *Planctomycetacia* (5.7%) and *Sphingobacteria* (3%) dominated the data set in all the samples. Although the class *Verrucomicrobiae* constituted 6.6% of all sequences, 88.5% of the sequences were from the river sample where the class represented

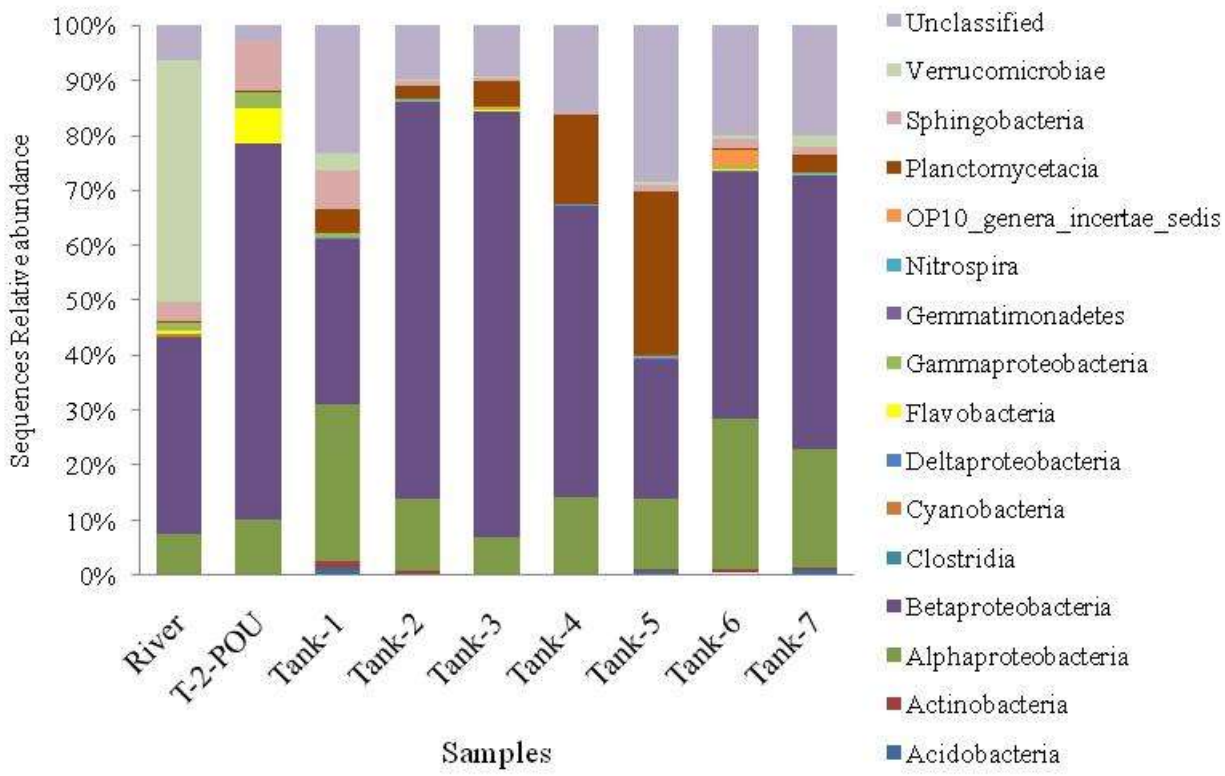

Figure 3: Average relative abundance (%) of the reads classified into the predominant microbial classes of individual samples.

43.7% of the observed sequences in the sample. The class *Flavobacteria* was detected in high abundance in the kitchen sample (6.5%), although it was also detected in river, tank 1 and tank 6 samples where it represented less than 1% of the observed sequences. Other classes with similar distribution were Clostridia that was only detected in tank 1 at 0.13%; *Cyanobacteria* in river water (0.7%); *Deltaproteobacteria* in tank 1(0.13%), tank 5 (0.35%) and tank 6 (0.17%); *Gemmatimonadetes* only in tank 5 (0.12%); *Nitrospira* in tank 2 (0.34%), tank 4 (0.08%) and tank 7 (0.1%) and *Armatimonadetes* in tank 3 (0.13%) and tank 6 (3.4%).

Diversity at order level

At order level 67.9% of sequences were classified into twenty-four orders. However, of the 67.9% classified sequences, 91.2% belonged to five orders; *Burkholderiales* (64.2%), *Verrucomicrobiales* (9.7%), *Planctomycetales* (8.4%), *Sphingobacteriales* (4.4%), *Sphingomonadales* (4.6%), *Rhizobiales* (1.6%) (Figure 4). Other orders that had abundance frequencies >1% were *Rhizobiales* (1.6%), *Rhodobacterales* (1.3%) and *Flavobacteriales* (1.1%). There were large variations in the relative abundances of a number of orders between

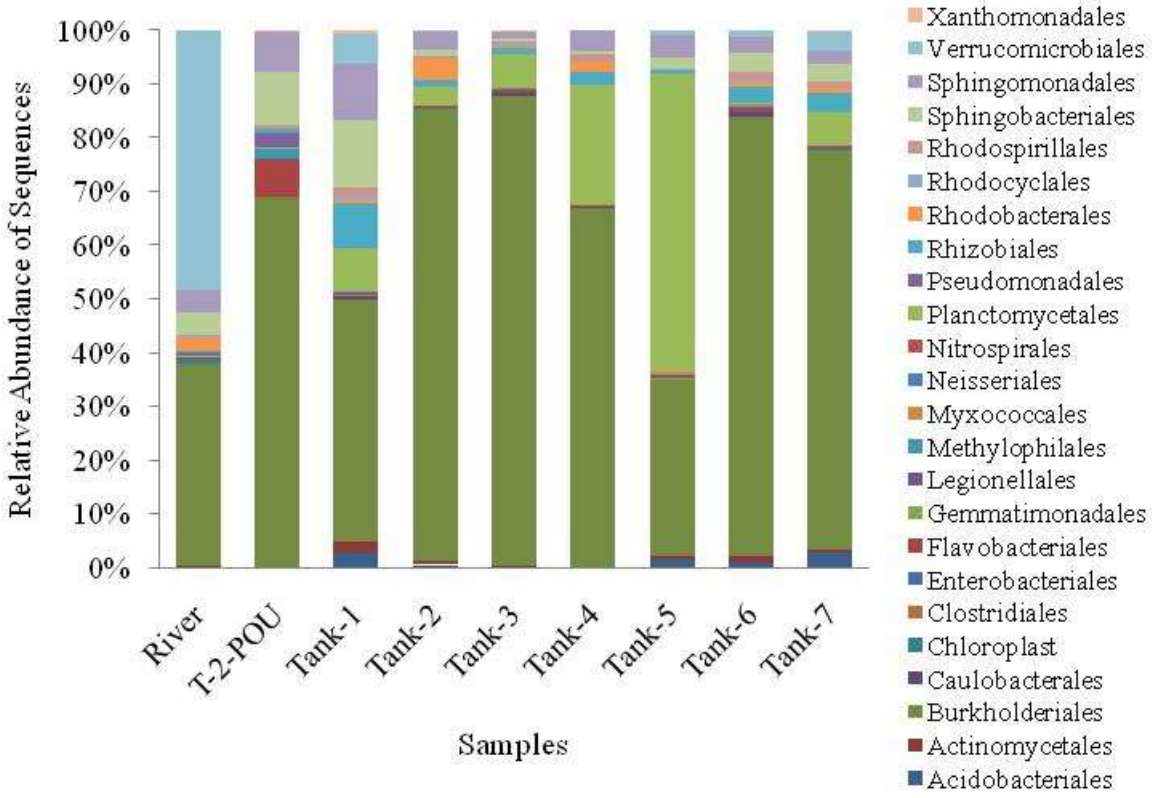

Figure 4: Average relative abundance (%) of the reads classified into the predominant microbial classes of individual samples.

individual samples. The order *verrucomicrobiales* was mostly abundant in river water sample (48.22), *Flavobacteriales* in kitchen sample (7%), *Sphingobacteriales* in kitchen (9.81) and tank 1 (12.5%) samples, *Plactomycetales* in tank 4 (22.2%) and 5 (55.5%). Wide relative abundance variation of particular orders in individual samples were observed for the orders *Burkholderiales* (32.8-87.3%), *Verrucomicrobiales* (0.16-48.2%) and *Planctomycetales* (0.07-55.5%). Orders that were not widely distributed among the samples included Chloroplasts that were only detected in river water (0.8%), *Clostridiales* in kitchen sample (0.24%), *Neisseriales* in tank 1 (0.12%), *Pseudomonadales* in river (0.94%) and kitchen (2.8%) samples.

Diversity at family level

At family level 41.6% of sequences were classified into thirty-seven families of which the most abundant were *Comamonadaceae* (33.1%), *Oxalobacteraceae* (16.6%),

Planctomycetaceae (13.7%), *Sphingomonadaceae* (7.4%) (Figure 5). Other abundant families included *Burkholderiaceae, Crenotrichaceae, Flexibacteraceae, Rhodobacteraceae,*

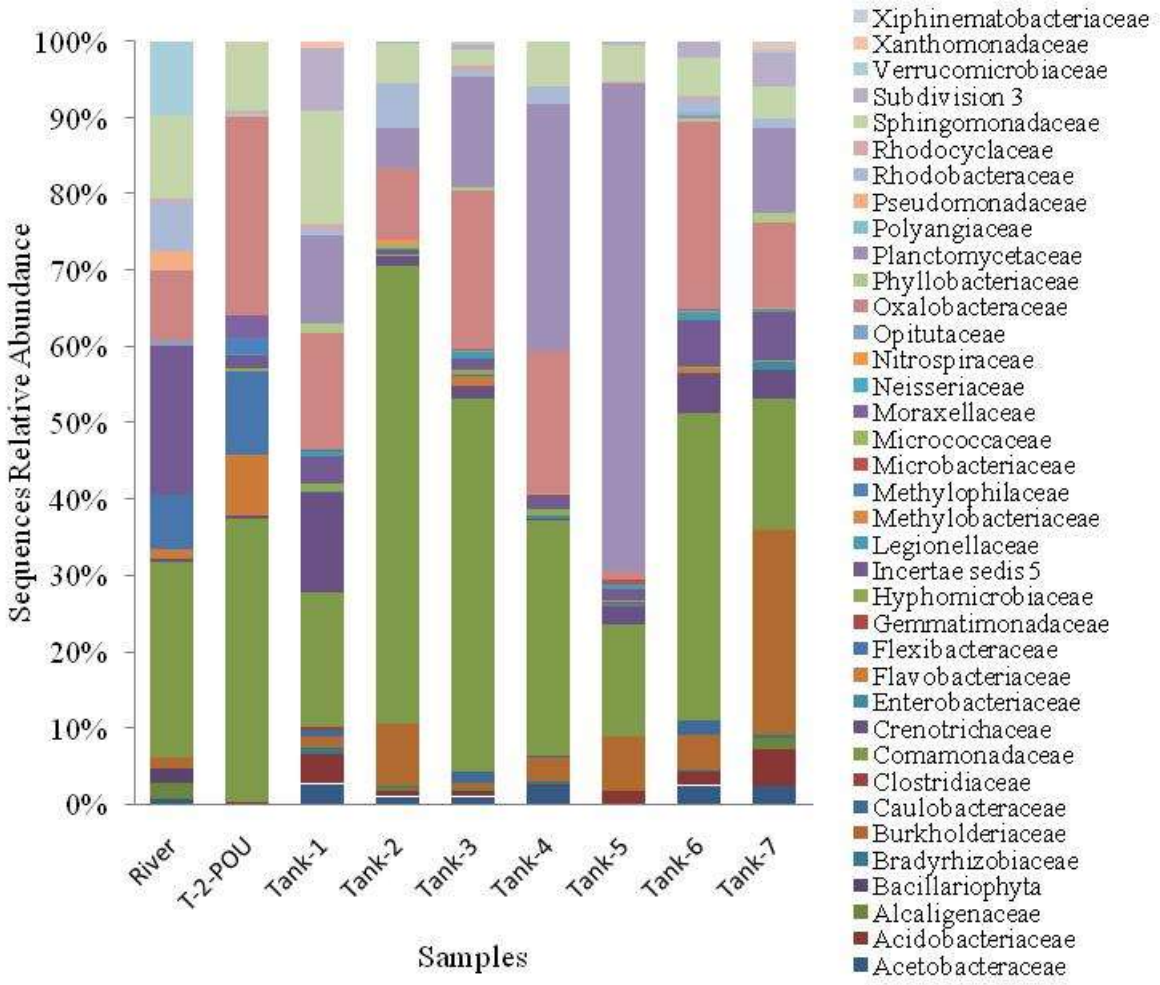

Figure 5: Relative abundance (%) of sequences classified into respective bacteria families of the individual samples.

Flavobacteriaceae, Acetobacteraceae, Acidobacteriaceae and *Verrucomicrobiaceae,* of which had relative abundances >1%. Large variations in relative abundances were also observed between samples. These were observed in families that included *Comamonadaceae* (14.61-60.1%), *Burkholriaceae* (0.59-26.9%), *Plannctomycetaceae* (0.41-64%) and *Crenotrichaceae* (0.43-13.2%). Families that were unique to particular samples included *Opitutaceae* (0.54%), *Bacillariophyta* (2%) and *Verrucomicrobiaceae* (9.6%) in river sample; *Moraxellaceae* in kitchen sample (2.9%); *Microbacteriaceae* in tank 1 (0.48%) and tank 5 (0.76%); *Pseudomonadaceae* in river (2.4%) and kitchen samples (0.34%).

Genus level diversity

Uclust's Venn diagram analysis was used to identify the genera with shared OTUs among samples (3% cut-off) as shown in Figure 6. In the Venn diagram, seven tank water samples

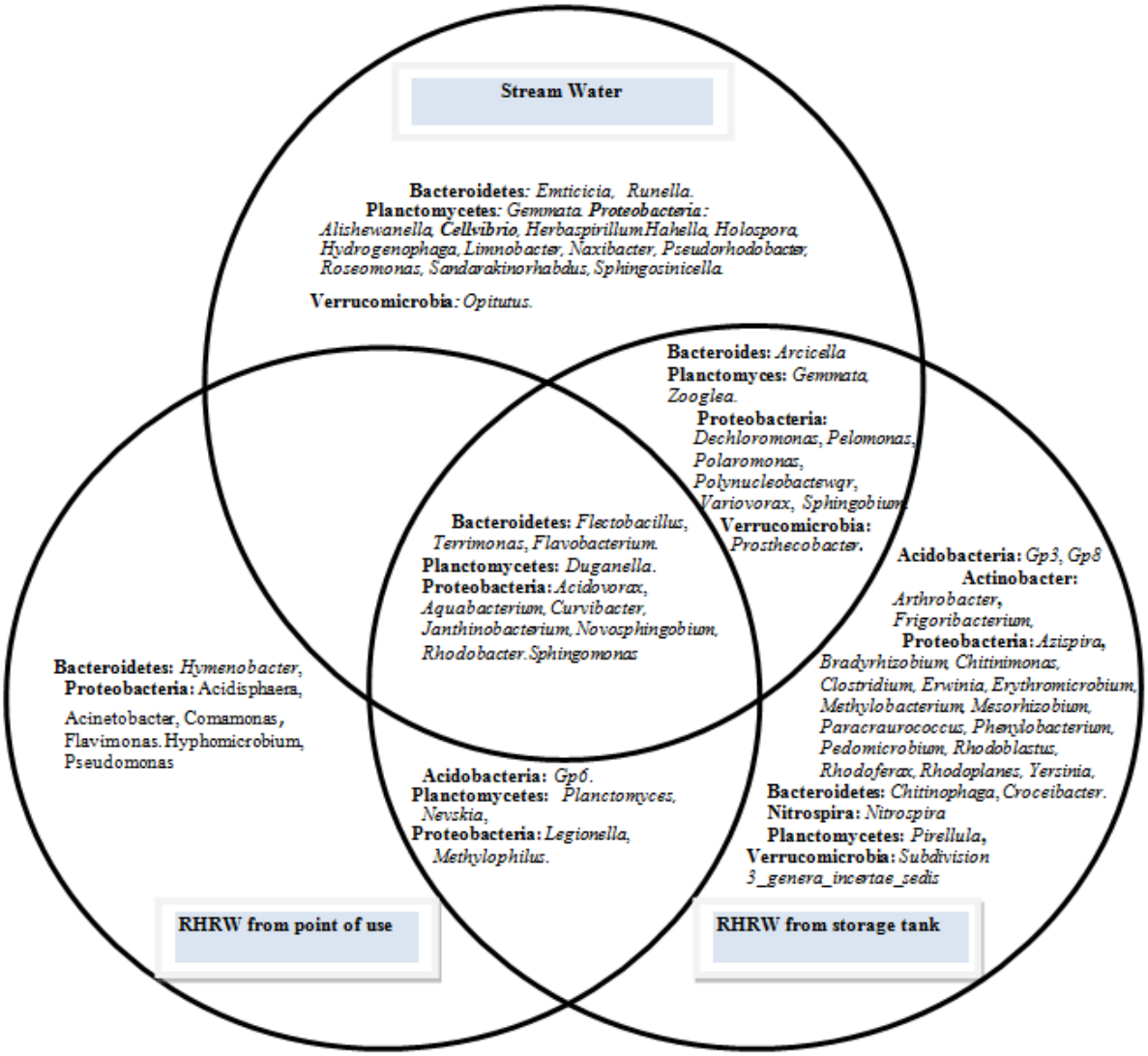

Figure 6: Schematic drawing showing the detected bacterial genera found in the three different groups of water samples (Tank, River and point of use (kitchen)) in Luthegele village. Overlaps between the groups are indicated by the arrangement of the circles.

were pooled to represent tank water whereas individual samples were used for river and kitchen water samples. Only eleven genera were shared among the three sample groups while ten were shared between tank water and river water samples; five between tank water and kitchen water samples and no unique genera was shared between river water and kitchen water. The most abundant genera in both tank and kitchen water belonged to the phylum *Proteobacteria* whereas in river water it was genera of the phylum *Planctomyces*. At genus

level, differences between the communities of various water sources were detected (Figure 7). Relative abundance of sequences in samples that could be classified into genera ranged from 12-54.2%, with the highest classified genera observed in POU samples. In total, 60 genera were detected from the total amplicon sequences library. Only a minor part of sequences belonging to members of the genera *Flectobacillus, Terrimonas, Flavobacterium, Duganella, Acidovorax, Aquabacterium, Curvibacter, Janthinobacterium, Novosphingobium, Rhodobacter and Sphingomonas* were retrieved from the three sample groups. Figure 7 also indicates that tank water samples contained a greater variety of bacteria (50 genera) than the river water sample (38) genera) and that from the kitchen water (22 genera). An overlap between the detected genera among the control samples and RHRW were also observed; the highest was found between river and tank water (10 genera), followed by tank and kitchen water (5 genera).

A comparison of the relative abundances of the classified bacterial genus's in the water samples confirmed differences between the samples (Figure 7). Most notably, *Curvibacter* was the most abundant genera on all sampling sites. In river water (Figure 7A), the most frequently present genus was *Arcinella* (12.6%) and *Curvibacter* (19.9%) among the other commonly found genera *Acidovorax, Aquabacterium, Cellvibrio, Janthinobacterium, Novosphingobium, Prosthecobacter* and *Pseudorhodobacter*. The dominant genera in kitchen water (Figure 7B) were Curvibacter (23%), *Janthinoacterium* (19%), *Acidovorax* (13.2%) *Flectobacillus* (15.2%), and *Novosphingobium* (12.1%). In tank water (Figure 7C) *Curvibacter* (38.7%), *Gemmata* (12.9%) *Polynucleobacter* (11%) and *Terrimonas* (6%) were the most abundant genus. Uniquely distributed genuses were *Acinebacter, Duganella, Methylophilus* detected only in kitchen water, *Cryobacterium* in tank 5, *Gp6 (uncultured bacterium), Nevskia* and *Planctomyces* in tank 1, *Hydrogenophaga, Cellvibrio, Opitutus, Prosthecobacter, Pseudohodobacter, Roseomonas* and *Sphingobium* in river water and *Rhodoplanes* in tank 7.

UniFrac analysis

Taxonomic categories dependent statistical tests often fail to detect differences in community level diversity (MacLean et al. 2009). Thus, approaches that are independent of OTU

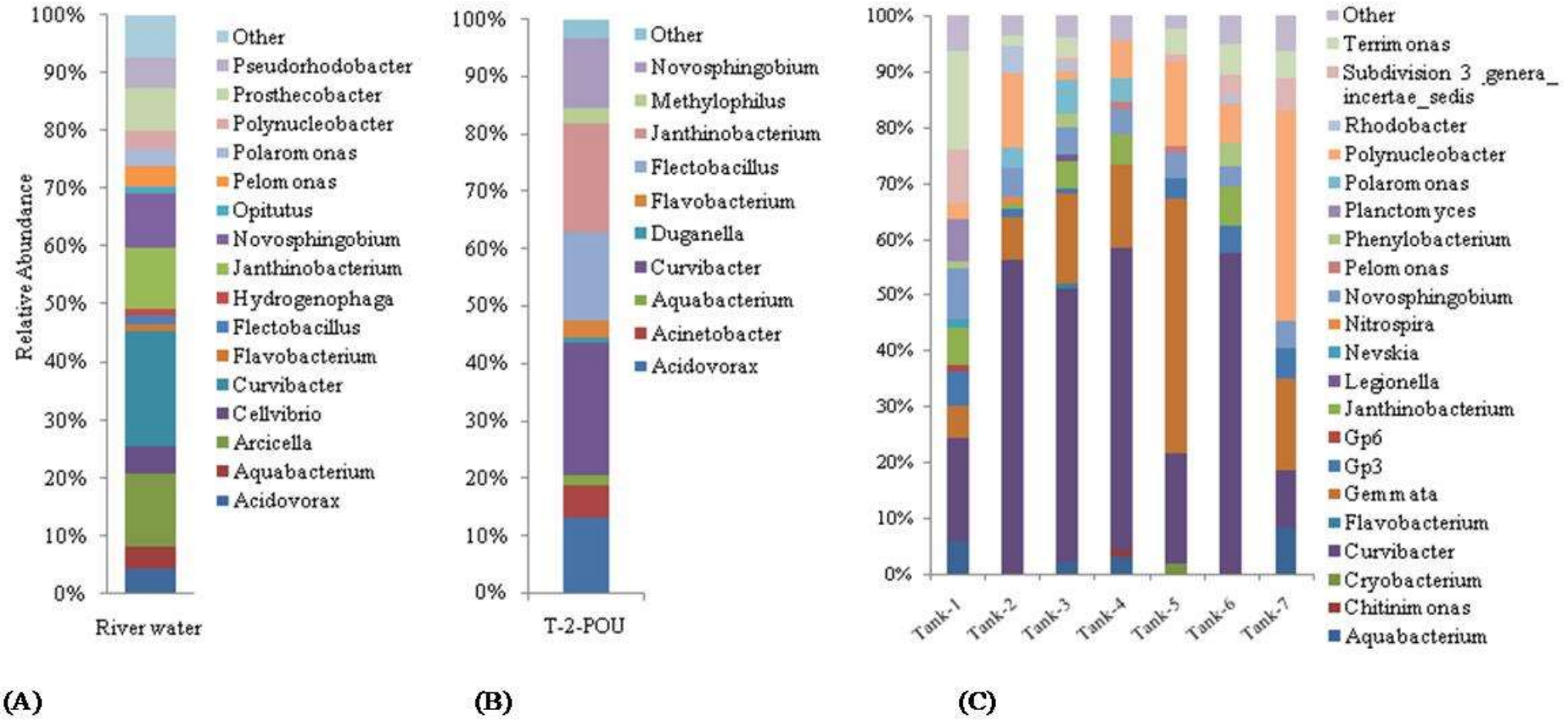

Figure 7 Bacterial communities of river water (A), kitchen water (B) and tank water (C). Relative clone composition of genera was determined by pyrosequencing of 16S rRNA from metagenomic DNA extracted from the water collected from Luthengele village in the Eastern Cape Province. The identification of the closest strain based on 16S rRNA sequence similarity was achieved using the web server SnoWMAn 1.15 (http://snowman.genome.tugraz.at). Phylogenetic groups accounting for 1% of all quality sequences are summarised in the artificial group "Other". Multi-coloured charts at the legend are shown for each genus and sample correspondingly.

assignments have been developed for comparing microbial communities (Lozupone et al. 2007). River, kitchen and seven RHRW samples were tested for community level differences in diversity with UniFrac Distance. The phylogenetic diversity comparison within groups by UniFrac Distance is independent of taxonomic classification (Lozupone et al. 2010). The results for the analysis were visualised through PCoA. Differences between samples obtained with Jacknife weighted and unweighted UniFrac distances showed marked clustering differences (Figure 8). River and kitchen water samples were clustered away from each other and from the seven tank water samples. Five of the tank water samples, (tanks 1, 2, 5, 6 and 7) clustered closely together as were samples for tanks 3 and 4.

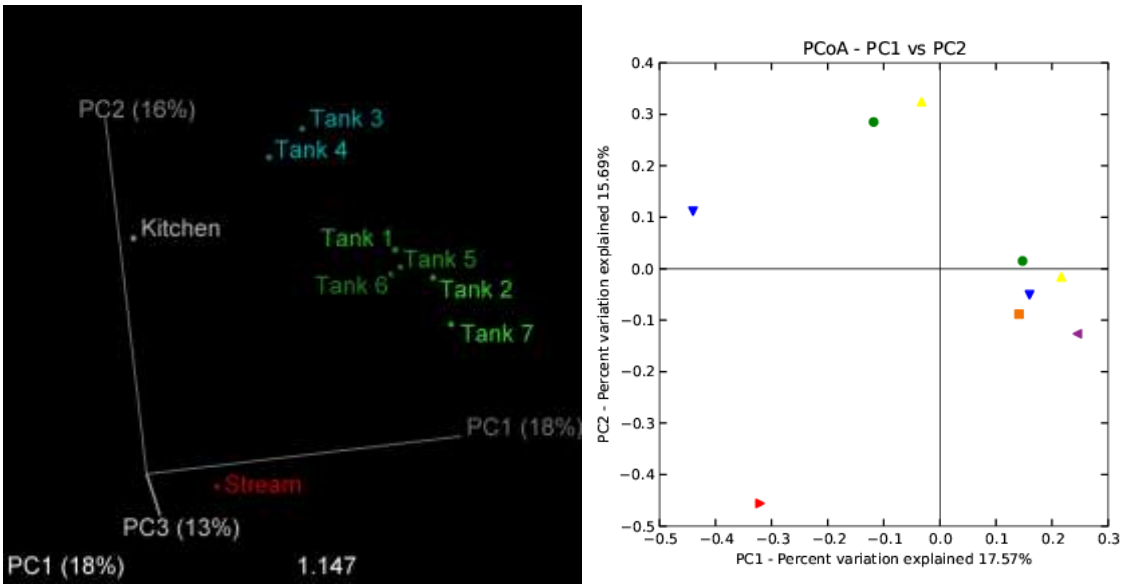

Figure 8: Principal coordinate analysis of unweighted UniFrac distances. PCoA plots were made using jackknifed UniFrac distances in QIIME, (A) PC1 versus PC2, (B) PC1 versus PC3, (C) PC2 versus PC3.

Detection of pathogenic signatures

A total of eight genera that are known to contain pathogenic bacteria were considered (Table 2). The 16S rRNA gene sequences from these genera were retrieved and the closest related species identified through Blastn from the NCBI database. These were then compared through a distance matrix constructed by the Jukes-Cantor algorithm. The sequences that were within a 0.03 Jukes-Cantor distance of the selected known species were considered correctly classified and are presented in Table 2.

Table 2: Probable species* of the detected genera known to contain pathogenic species

Genera	Species
Serratia	*nematodiphila, ureilytica*
Tatlockia,	*micdadei*
Chromobacterium	*aquaticum, haemolyticum, violaceum, subtsugae*
Clostridium	*algidicarnis, putrefaciens, sartagoforme, cellulovorans*
Pseudomonas	*brassicacearum_subsp. denitrificans, entomophila, putida, nitroreducens, fluorescens*
Legionella	*pneumophila_subsp._pascullei, pneumophila_subsp._pneumophila, cincinnatiensis, longbeachae, fairfieldensis, adelaidensis, rubrilucens, taurinensis, gresilensis, uncultured_Legionella_sp.*
Yersinia	*bercovieri, pestis, mollaretii, frederiksenii, ruckeri, massiliensis*
Acinetobacter	*oleivorans, calcoaceticus, haemolyticus, junii, parvus, gyllenbergii*

*Analyses were conducted using the Jukes-Cantor model (Jukes and Cantor 1969) and evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011).

The similarity between our sequences and those of known related bacteria are shown in the Neighbour-Joining tree dendrogram that was constructed with evolutionary distances computed using the Jukes-Cantor method (Figure 9). However, we could not assign the identity of our sequences to their respective unique species. A common feature of the dendrogram was that a particular sequence was grouped together with two or more different species. In the *Acinetobacter* cluster, *Acinetobacter parvus and Arcinetobacter gyllenbergii* were grouped at the same level with six *Acinebacter* sequences. Similarly in the *Pseudomonas* group the species *P. denitrificans, P. entomophila, P. putida, P. Fluorescens and P nirtoreducens* were grouped together with three sequences. This trend was similarly observed in the rest of the cluster groups for the detected signatures of potentially pathogenic bacteria.

Legionella signatures were the most prevalent and were detected in six of the nine samples. The six sequences of the *Acinetobacter* and three of *Pseudomonas* were only detected in

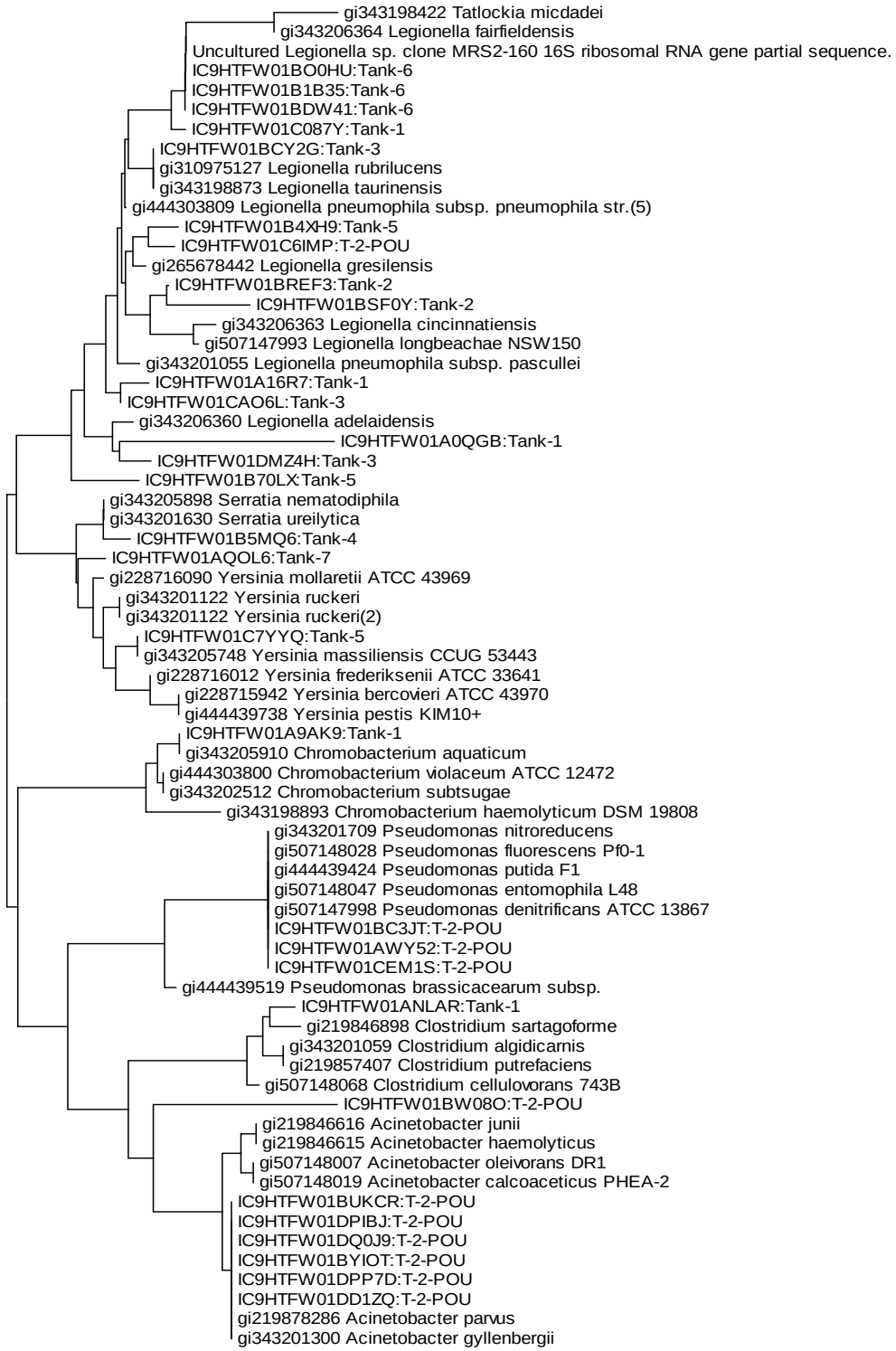

Figure 9: Evolutionary relationships of taxa, The evolutionary history was inferred using the Neighbour-Joining method (Saitou & Nei 1987). The evolutionary distances were computed using the Jukes-Cantor method (Jukes & Cantor 1969) and are in the units of the number of base substitutions per site.. Evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011)

kitchen water, *Chlostridia* (1 sequence) and *Chromobacterium* (1 sequence) in tank 1, *Yersinia* (2) in tank 1 and tank 5 and *Serratia* in tank 4 and tank 7. Of the species that were identified to be closely related to the sequences of our pathogenic signatures, only the species *Legionella pneumophila* is included in the U.S. EPA bacteria of concern in water (US EPA 2009).

Discussion

Little information exists about microbial communities in RHRW. In this study we used barcoded pyrosequencing of the V1-V3 hypervariable regions of the 16S RNA to investigate microbial communities in RHRW used for domestic purposes in Luthengele village, Eastern Cape Province of South Africa. Households in this village solely depend on RHRW and stream water sources. During the wet season, households depend on RHRW and in drier months, it is supplemented with stream water. However, the microbial safety of RHRW has been disputed by a number of researches that reported the presence of pathogenic bacteria and parasites (Lye 2009; Ahmed et al. 2010b). Furthermore the prevalence of domestic animals in this village is another risk factor given that domestic animals have been reported to harbour pathogenic bacteria and parasites which they spread through their faecal material (Ashbolt 2004).

The results of our study demonstrated diverse bacterial communities in RHRW, river and kitchen water samples. However, a prominent feature of our analysis was the presence of unclassified bacteria. These findings suggest for the existence of diverse bacterial that presumably characterise the microbial community structure in the water samples (Vaz-Moreira et al. 2011).

Of the eleven different phyla detected in water samples, phylotypes belonging to the phylums *Proteobacteria*, *Verrucomicrobia*, *Plancomyces* and *Bacteroidetes* dominated the microbial communities. Members of these phyla are common inhabitants of freshwater and both *Proteobacteria* and *Bacteroidetes* have been reported using culture-dependent, high-throughput cultivation (Gich et al. 2005) or culture-independent, 16S rRNA based clonal analyses (Matcher et al. 2011; Vaz-Moreira et al. 2011; Cleary et al. 2012; Steelman et al.

2012) and metagenomic library and FISH analysis methods (Cottrell et al. 2005). However, the phylums, *Verrucomicrobia* and *Plancomyces* have only been reported through culture-independent methods. *Firmicutes, Actinobacteria, Acidobacteria* and *Armatimonadetes* were found in lower abundances as previously pointed out in other freshwater studies (Hong et al. 2010; Matcher et al. 2011; Vaz-Moreira et al. 2011). At the family and genus taxonomic level, some families that included, *Comamonadaceae, Bacillariophyta, Burkholriaceae, Crenotrichaceae, Pseudomonadaceae* and *Legionellaceae* and the genera *Flavobacterium, Chitinophaga, Rhodoferax* and *Curvibacter* were previously observed in freshwater using 16S rRNA clone libraries (Zwart et al. 2002; Matcher et al. 2011; Vaz-Moreira et al. 2011; Cleary et al. 2012; Steelman et al. 2012), reverse line blot hybridisation (Lindström et al. 2005) or metagenomic and FISH analysis (Cottrell et al. 2005).

Bacteroidetes are a very diverse heterotrophic bacterial division with aerobic members being able to degrade biopolymers that could be present in the high-molecular weight fraction of the dissolved organic material such as cellulose, chitin and proteins (Kirchman, 2002). At the genus taxonomic level the presence of large proportions of phylotypes belonging to the genera *Acidovorax, Curvibacter* and *Gemmata*, and the family *Commamonadaceae* in our study was similar to other findings in fresh water samples from around the world (Pinto et al. 2012; Lautenschlager et al. 2013; Navarro-noya et al. 2013).

In comparison, *Verrucomicrobia* were mostly dominant in river water (44.5%) and were detected at less than 2% in RHRW samples. The phylum *Verrucomicrobia* is ubiquitous in soil and water microbial communities where it sometimes can be detected in high abundance (Janssen 2006). However, little is known about the distribution and diversity of this phylum in freshwater ecosystems. The phylum is related to *Planctomycetes* and *Chlamydiales* and is morphologically diverse, including the presence of intracellular compartments. The known members of *Verrucomicrobia* nearly all grow chemoheterotrophically on many organic compounds and pose no known health risk to humans (Schlesner et al. 2006). *Planctomyces* were prevalent in six of the seven tanks and are composed of free-living and ubiquitous species in a wide range of terrestrial and aquatic environments. Information of this group is limited owing to relatively few species that have been obtained in pure culture and sequenced (Ward et al. 2006).

Several researchers have reported on the abundance of *Betaproteobacteria* and *Alphaproteobacteria* in freshwater systems (Chen et al. 2004; Eichler et al. 2006). The dominance of *Betaproteobacteria* and *Alphaproteobacteria* demonstrated the likelihood that fresh water ecosystems provided a conducive environment for enriching members within *Betaproteobacteria* and *Alphaproteobacteria*. However, further research is needed to evaluate the factors responsible for the competitive advantage these two classes of the phylum *Proteobateria* have over other bacterial species.

Environmental influence on microbial diversity

The interaction between a variety of factors which include the availability of required nutrients and the mortality rates due to biological and/ or physical factors are responsible for the observed bacterial community composition in any given specific environment (Dolan 2005). Freshwater habitats comprise a specific bacterial community in which *Proteobacteria* (mostly *alpha-*, *beta-* and *gamma Proteobacteria*) are dominant, although *Actinobacteria*, *Bacteroidetes*, *Nitrospira*, *Chloroflexi*, *Cyanobacteria*, and *Verrucomicrobia* are also frequently present (Zwart et al. 2002). The dominance of *Proteobacteria* in drinking water has been demonstrated by a number of researchers (Zwart et al. 2002; Chen et al. 2004; Bricheux et al. 2013). However the presence of other bacteria phylum will depend on the immediate sources of contamination or the fresh water ecosystem (Eichler et al. 2006). River water has direct contact with soil, is exposed to light and has increased amounts of organic matter from rotting vegetation and faecal matter from domestic animals. Roof-harvested rainwater has no such direct contact with the environment besides roof runoff that flows into the tank. RHRW in tanks forms a closed ecosystem whose microbial ecology is a function of inputs from roof runoff and biofilm formation in the tank.

Roof-harvested rainwater is stored in tanks from which it is collected and stored in smaller containers in the kitchens prior to use. The process of collecting this water exposes it to potential contamination from handling and unclean containers. Although we used single samples to evaluate this, our findings showed that there were significant differences in the microbial communities between kitchen water and the source tank (tank 2). Relatively higher

abundance of the classes *Gammaproteobateria* and *Flavobacteria* were observed in kitchen water than anywhere among the samples including tank 2 from which the kitchen water was collected. However further studies that include a significantly higher number of samples are needed to confirm these findings and further explore the influence of water handling practices on RHRW microbiological quality.

Species richness

The OTU rarefaction plot failed to plateau at species or genus level, signifying lower bacterial richness and diversity despite a relatively high Shannon bacterial diversity index (5.24 to 7.5). This could be due to insufficient number of reads to capture all of the diversity within each sample, particularly for the low abundance OTUs. This demonstrates the need for more sampling events and generation of more sequences i.e. 5000 reads per sample than the 855-1739 reads per sample generated in this study.

The numbers of 16S rRNA gene sequences analysed (10956-18510 sequences) is significantly larger than the number of sequences employed in analysis that use conventional cloning and sequencing methods, of which sequences used in these analysis are generally less than 200. The increased number of sequences made the detection of more microorganisms possible (i.e. 482-740 OTUs at a 3% cut-off). These numbers are much greater than those based on clone libraries as reported in previous studies, of which less than 100 OTUs were identified from freshwater samples (Zwart et al. 2002; Eichler et al. 2006) and drinking water treatment systems (Martiny et al. 2005; Eichler et al. 2006; Poitelon et al. 2009). The observed species in this study (184.4-507.1) were even higher than those observed by Navarro-Noya et al. (2013) (27.9 and 239.4) in drinking water wells and by Hong et al. (2010) (133 and 208 OTUs) from biofilm in two water meters investigated.

Detection of pathogenic signatures

Generally, it is impossible to predict the pathogenicity of a strain based on 16S rRNA sequences, although signatures of potentially pathogenic bacteria can be detected. However, the use of 16S rDNA method for species identification is restricted to bacterial pathogens and

hence, excludes fungi and human parasites. Of the species that were closely related to our detected pathogenic signatures for the genera, *Acinetobacter, Clostridium, Legionella, Pseudomonas, Seratia* and *Yersinia*, only the species *Legionella pneumophila* is included in the U.S. EPA bacteria of concern in water (US EPA 2009). *Legionella* species occur naturally in environmental water, including groundwater (Costa et al. 2005), surface water (Palmer et al. 1993) and have also been reported in RHRW stored in tanks (Ahmed et al. 2008). *Legionella* spp. are ubiquitous in nature and can survive under varied water conditions, and their survival is due in part, attributed to their interactions with other members of the heterotrophic flora. *Legionella* spp. can form symbiotic relationships with other bacteria, such as *Alcaligenes, Acinetobacter, Flavobacterium* and *Pseudomonas* and is thought to be critical for their survival and proliferation in water (Stout et al. 1985). Furthermore, some naturally occurring protozoa in water, such as *Acanthamoeba castellanii* and *Hartmanella* spp., can harbour *Legionella* organisms, shielding them from environmental stresses and providing a conducive environment for their proliferation (Stout et al. 1985; Rowbotham 1986). This may explain why *Legionella* was widely distributed in our samples. Furthermore it can be hypothesised that the *Legionella* spp. may be proliferating in the tank biofilms as has been previously reported (Murga et al. 2001).

Approximately half of the *Legionella* species have been associated with disease in humans, with *Legionella pneumophila* infection being responsible for the majority of illnesses. Other than *L. pneumophila*, human illnesses have also resulted from infection with *L. micdadei*, *L. bozemanii, L. longbeachae,* and *L. dumoffi* and many other *Legionella* spp. have been implicated (Fields et al. 2002). Pneumonia is the major clinical manifestation of infection by *Legionella* species, although nonpneumonic legionellosis (Pontiac fever) and extrapulmonary infection may occur (Muder and Yu 2002). The majority of non-pneumophila *Legionella* confirmed infections have been reported in immunosuppressed patients.

Pathogenic bacteria of concern that have previously been identified in RHRW include *Clostridium, Bacteroides, Campylobacter, Legionella, Aeromonas, Salmonella* and *Yersinia* (Ahmed et al. 2008; Ahmed et al. 2010b). While we were able to detect sequences belonging to pathogenic genera including *Acinetobacter, Clostridium, Flavobacterium, Legionella, Methylobacterium, Pseudomonas, Seratia and Yersinia.* their values were low (1 in most

cases) and more sequencing depth is required to either exclude them or more confidently describe their presence. Previous estimates of risk from *Salmonella* spp. due to liquid ingestion via drinking suggest that a viable pathogen level 6.9 x 10^1-3.8x 10^2 per single event say one litre consumption may lead to a 1 in 10,000 risk, (Ahmed et al. 2010b). A10^2 value, when compared to, say 10^{10} total cells per litter of water requires 10^8 sequences to identify a pathogenic sequence at this enrichment (Bibby et al. 2010). Given the rate at which sequencing technology has progressed, pyrosequencing may be able to provide this level of sensitivity in the near future (Metzker 2009). In order to investigate the prevalence of rare pathogenic signatures and to fully describe their diversity, multiple genus level PCR primers may need to be employed. Moreover, recent studies that tracked the total concentrations of pathogens in RHRW suggest that qPCR-based detection may be a reasonable indicator of pathogen load (Ahmed et al. 2011).

Although, the use of pyrosequencing to construct sequence libraries has enabled new and important insights into pathogen loads and diversity in the environment, the use of Jukes-Cantor's 0.03 distances identifying a species takes on added importance. Although the maximum of 500 base length generated by the current sequencing technologies is a significant improvement over 100 or 250 bases previously used, it may still not allow for definitive, phylogenetic identification of pathogens. Although the 0.03 Jukes-Cantor distance is well accepted as operational definition) it may not define a true pathogen. as previously described by Bibby et al. (2010). Similarly in our findings on the Legionella spp. we could not confirm the identity of some sequences to belong to a specific species as multiple species were clustered together at distances lower than 0.03 Jukes-Cantor distance (Figure 9). Without additional information, it cannot be determined which species the unknown sequences belong to and whether they are strains or variants of the species not yet identified. In absence of sequences describing all pathogens associated with potentially pathogenic bacteria and their variants, the sequencing of environmental sample based on universal primers does not as yet have the resolving power to recognise the significance of sequences closely related to known species and those outside of accepted species level distance. This power will improve as research continues to populate databases with sequences of known or suspected potentially pathogenic bacteria.

Conclusion and recommendations

Our findings demonstrated the potential of pyrosequencing as a method to define sequences of potentially pathogenic species, and highlighted the need to determine the necessary sequencing depth required to describe microbial risk levels. However, the limitations to this approach include, the potential uncertainty in accurate pathogen identification using the 16s rRNA gene and the need for greater sequencing depth to attain a level where primary pathogens in water samples are in abundance. In the future, pyrosequencing analysis of environmental samples to detect the presence of pathogens should build upon existing studies to decide on sequencing targets, sequencing depth, and bioinformatic techniques. Future, studies on the microbial quality of RHRW should include qPCR quantification of potential pathogens to supplement pyrosequencing data and be able to estimate actual level of microbiological risk. This should include the use of multiple samples from both RHRW tanks alternative water sources used by the households.

Acknowledgements

This study was undertaken as part of a Water Research Commission (WRC) unsolicited project: "*Evaluation of the risks associated with the use of rainwater harvested from roofs, for domestic use and, homestead food gardens; and groundwater for domestic use and livestock watering*" (WRC Project No K5/2175, Water Research Commission, 2013).

References

Ahmed, W., Gardner, T., & Toze, S. (2011). Microbiological quality of roof-harvested rainwater and health risks: A review. Journal of Environmental Quality, 40(1), 13-21.

Ahmed, W., Goonetilleke, A., & Gardner, T. (2010a). Implications of faecal indicator bacteria for the microbiological assessment of roof-harvested rainwater quality in southeast Queensland, Australia. Canadian Journal of Microbiology, 56(6), 471-479.

Ahmed, W., Huygens, F., Goonetilleke, A., & Gardner, T. (2008). Real-Time PCR detection of pathogenic microorganisms in roof-harvested rainwater in Southeast Queensland, Australia. Applied and Environmental Microbiology, 74(17), 5490-5496.

Ahmed, W., Sidhu, J. P. S., Toze, S., & Hodgers, L. (2012a). Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland. Applied and Environmental Microbiology, 78(1), 219-226.

Ahmed, W., Vieritz, A., Goonetilleke, A., & Gardner, T. (2010b). Health risk from the use of roof-harvested rainwater in Southeast Queensland, Australia, as potable or nonpotable water, determined using quantitative microbial risk assessment. Applied and Environmental Microbiology, 76(22), 7382-7391.

Ahmed, W, J., Sidhu, P.S., & Toze. S. (2012b). An attempt to identify the likely sources of Escherichia coli harboring toxin genes in rainwater tanks. Environmental Science & Technology, 46(9), 5193-5197.

Ashbolt, N. J. (2004). Microbial contamination of drinking water and disease outcomes in developing regions. Toxicology, 198(1), 229-238.

Berg, G., Eberl, L., & Hartmann, A. (2005). The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. Environmental Microbiology, 7(11), 1673-1685.

Bibby, K., Viau, E., & Peccia, J. (2010). Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. Water Research, 44(14), 4252-4260.

Bricheux, G., Morin, L., Moal, G., Coffe, G., Balestrino, D., Charbonnel, N., et al. (2013). Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. Microbiology Open, 2(3), 402-414.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335-336.

Chen, C. L., Liu, W. T., Chong, M. L., Wong, M. T., Ong, S. L., Seah, H., et al. (2004). Community structure of microbial biofilms associated with membrane-based water purification processes as revealed using a polyphasic approach. Applied Microbiology and Biotechnology, 63(4), 466-473.

Cleary, D. F. R., Smalla, K., Mendonc, L. C. S., & Gomes, N. C. M. (2012). Assessment of variation in bacterial composition among microhabitats in a mangrove environment using DGGE fingerprints and barcoded pyrosequencing. PloS One, 7(1), e29380.

Costa, J., Tiago, I., da Costa, M. S., & Veríssimo, A. (2005). Presence and persistence of *Legionella* spp. in ground-water. Applied and Environmental Microbiology, 71(2), 663-671.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science, 326(5960), 1694-1697.

Cottrell, M. T., Waidner, L. A., Yu, L., & Kirchman, D. L. (2005). Bacterial diversity of metagenomic and PCR libraries from the Delaware River. Environmental Microbiology, 7(12), 1883-1895.

DeSantis, T. Z., Brodie, E. L., Moberg, J. P., Zubieta, I. X., Piceno, Y. M., & Andersen, G. L. (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microbial Ecology, 53(3), 371-383.

Dolan, J. R. (2005). Biogeography of aquatic microbes. Aquatic Microbial Ecology, 41, 39-48.

Eichler, S., Christen, R., Höltje, C., Westphal, P., Bötel, J., Brettar, I., et al. (2006). Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA-and DNA-based 16S rRNA gene fingerprinting. Applied and Environmental Microbiology, 72(3), 1858-1872.

Fields, B. S., Benson, R. F., & Besser, R. E. (2002). *Legionella* and Legionnaires' disease: 25 years of investigation. Clinical Microbiology Reviews, 15(3), 506-526.

Gich, F., Schubert, K., Bruns, A., Hoffelner, H., & Overmann, J. (2005). Specific detection, isolation, and characterization of selected, previously uncultured members of the freshwater bacterio-plankton community. Applied and Environmental Microbiology, 71(10), 5908-5919.

Hong, P.Y., Hwang, C., Ling, F., Andersen, G. L., Le Chevallier, M. W., Liu, W., et al. (2010). Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. Applied and Environmental Microbiology, 76(16), 5631-5635.

Hur, I., & Chun, J. (2004). A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. Journal of Microbiology, 42(1), 9-13.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Applied and Environmental Microbiology, 72(3), 1719-1728.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules, Munro HN, Mammalian protein metabolism (pp. 21-132). New York: Academic press.

Kahinda, J. M., Taigbenu, A. E., & Boroto, R. J. (2010). Domestic rainwater harvesting as an adaptation measure to climate change in South Africa. Physics and Chemistry of the Earth Parts ABC, 35, 742-751.

Kaushik, R., Balasubramanian, R., & de la Cruz, A. A. (2012). Influence of Air Quality on the Composition of Microbial Pathogens in Fresh Rainwater. Applied and Environmental Microbiology, 78(8), 2813-2818.

Kaushik, R., Balasubramanian, R., & Dunstan, H. (2014). Microbial Quality and Phylogenetic Diversity of Fresh Rainwater and Tropical Freshwater Reservoir. PLoS One, 9(6), e100737.

Kirchman, D. L. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. FEMS Microbial Ecology, 39(2), 91-100.

Lai, C.-C., Cheng, A., Liu, W.-L., Tan, C.-K., Huang, Y.-T., Chung, K.-P., et al. (2011). Infections caused by unusual *Methylobacterium* species. Journal of Clinical Microbiology, 49(9), 3329-3331.

Lautenschlager, K., Hwang, C., Liu, W., Boon, N., Ko, O., Vrouwenvelder, H., et al. (2013). A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. Water Research, 47, 3015-3025.

Lindström, E. S., Kamst-Van Agterveld, M. P., & Zwart, G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. Applied and Environmental Microbiology, 71(12), 8201-8206.

Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental microbiology, 73(5), 1576-1585.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2010). UniFrac: An effective distance metric for microbial community comparison. The ISME Journal, 5(2), 169-172.

Luxmy, B., Nakajima, F., & Yamamoto, K. (2000). Analysis of bacterial community in membrane-separation bioreactors by fluorescent in situ hybridization (FISH) and denaturing gradient gel-electrophoresis (DGGE) techniques. Water Science & Technology, 41(10-11), 259-268.

Lye, D. J. (2009). Rooftop runoff as a source of contamination : A review. Science of the Total Environment, 407(21), 5429-5434.

MacLean, D., Jones, J. D. G., & Studholme, D. J. (2009). Application of next-generation' sequencing technologies to microbial genetics. Nature Reviews Microbiology, 7(4), 287-296.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in micro-fabricated high-density picolitre reactors. Nature, 437(7057), 376-380.

Martiny, A. C., Albrechtsen, H.-J., Arvin, E., & Molin, S. (2005). Identification of bacteria in biofilm and bulk water samples from a non-chlorinated model drinking water distribution system: Detection of a large nitrite-oxidizing population associated with Nitrospira spp. Applied and Environmental Microbiology, 71(12), 8611-8617.

Matcher, G. F., Dorrington, R. A., Henninger, T. O., & Froneman, P. W. (2011). Insights into the bacterial diversity in a freshwater-deprived permanently open Eastern Cape estuary, using 16S rRNA pyrosequencing analysis. Water SA, 37(3), 381-390.

Meays, C. L., Broersma, K., Nordin, R., & Mazumder, A. (2004). Source tracking faecal bacteria in water: A critical review of current methods. Journal of Environmental Management, 73(1), 71-79.

Metzker, M. L. (2009). Sequencing technologies: The next generation. Nature Reviews Genetics, 11(1), 31-46.

Muder, R. R., & Yu, V. L. (2002). Infection due to Legionella species other than *L. pneumophila*. Clinical Infectious Diseases, Clinic990-998.

Murga, R., Forster, T. S., Brown, E., Pruckler, J. M., Fields, B. S., & Donlan, R. M. (2001). Role of biofilms in the survival of *Legionella pneumophila* in a model potable-water system. Microbiology, 147(11), 3121-3126.

Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology, 59(3), 695-700.

Navarro-Noya, Y. E., Suárez-Arriaga, M. C., Rojas-Valdes, A., Montoya-Ciriaco, N. M., Gómez-Acata, S., Fernández-Luqueño, F., & Dendooven, L. (2013). Pyrosequencing Analysis of the Bacterial Community in Drinking Water Wells. Microbial Ecology, 66(1), 19-29.

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews, 75(1), 14-49.

Nocker, A., Burr, M., & Camper, A. K. (2007). Genotypic microbial community profiling: A critical technical review. Microbial Ecology, 54(2), 276-289.

Orgiazzi, A., Bianciotto, V., Bonfante, P., Daghino, S., Ghignone, S., Vizzini, A., et al. (2013). 454 Pyrosequencing analyses of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. Diversity, 5(1), 73-98.

Palmer, C. J., Tsai, Y. L., Paszko-Kolva, C., Mayer, C., & Sangermano, L. R. (1993). Detection of *Legionella* species in sewage and ocean water by polymerase chain reaction, direct fluorescent-antibody, and plate culture methods. Applied and Environmental Microbiology, 59(11), 3618-3624.

Paulse, A. N., Jackson, V. A., & Khan, W. (2009). Comparison of microbial contamination at various sites along the Plankenburg- and Diep Rivers, Western Cape, South Africa. Water SA, 35(4), 469-478.

Pinto, A. J., Xi, C., & Raskin, L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environmental Science & Technology, 46(16), 8851-8859.

Poitelon, J.-B., Joyeux, M., Welté, B., Duguet, J.-P., Prestel, E., Lespinet, O., et al. (2009). Assessment of phylogenetic diversity of bacterial microflora in drinking water using serial analysis of ribosomal sequence tags. Water Research, 43(17), 4197-4206.

Rasid, R. A. (2009). Biofilm and multimedia filtration for rainwater treatment. Journal of Sustainable Development, 2(1), 196-199.

Roesch, L. F. W., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., & Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME Journal, 1(4), 283-290.

Rowbotham, T. J. (1986). Current views on the relationships between amoebae, legionellae and man. Israel Journal of Medical Sciences, 22(9), 678.

Rusin, P. A., Rose, J. B., Haas, C. N., & Gerba, C. P. (1997). Risk assessment of opportunistic bacterial pathogens in drinking water. In Reviews of Environmental Contamination and Toxicology (pp. 57-83). New York: Springer.

Sanapareddy, N., Hamp, T. J., Gonzalez, L. C., Hilger, H. A., Fodor, A. A., & Clinton, S. M. (2009). Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Applied and Environmental Microbiology, 75(6), 1688-1696.

Schlesner, H., Jenkins, C., & Staley, J. T. (2006). The phylum *Verrucomicrobia*: A phylogenetically heterogeneous bacterial group. In The Prokaryotes (pp. 881-896). New York: Springer.

Shanks, O. C., White, K., Kelty, C. A., Sivaganesan, M., Blannon, J., Meckes, M., et al. (2010). Performance of PCR-based assays targeting *Bacteroidales* genetic markers of human faecal pollution in sewage and faecal samples. Environmental Science and Technology, 44, 6281.

Shi, P., Jia, S., Zhang, X. X., Zhang, T., Cheng, S., & Li, A. (2013). Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Research, 47(1), 111-120.

Sinclair, R., Boone, S. A., Greenberg, D., Keim, P., & Gerba, C. P. (2008). Persistence of category A select agents in the environment. Applied and Environmental Microbiology, 74(3), 555-563.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences, 103(32), 12115-12120.

Steelman, S. M., Chowdhary, B. P., Dowd, S., Suchodolski, J., & Jane, J. E. (2012). Pyrosequencing of 16S rRNA genes in faecal samples reveals high diversity of hindgut micro-flora in horses and potential links to chronic laminitis. BMC Veterinary Research, 8(231), 1-11.

Stocker, S., Snajder, R., Rainer, J., Trajanoski, S., Gorkiewicz, G., Trajanoski, Z., & Thallinger, G. G. (2011). SnoWMAn: High-throughput phylotyping, analysis and comparison of microbial communities. In Proceedings of the 110th ASM General Meeting 23-5-2010.

Stout, J. E., Yu, V. L., & Best, M. G. (1985). Ecology of *Legionella pneumophila* within water distribution systems. Applied and Environmental Microbiology, 49(1), 221-228.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10), 2731-2739.

US EPA. (2002). Health risks from microbial growth and biofilms in drinking water distribution systems. Washington, DC: U.S. Environmental Protection Agency, Office of Ground Water and Drinking Water Standards and Risk Management

US EPA. (2009). Drinking water contaminant candidate List 3. Federal Register, 74(194), 51850-51862.

Vaz-Moreira, I., Egas, C., Nunes, O. C., & Manaia, C. M. (2011). Culture-dependent and culture-independent diversity surveys target different bacteria: A case study in a freshwater sample. Antonie Van Leeuwenhoek, 100(2), 245-257.

Ward, N., Staley, J. T., Fuerst, J. A., Giovannoni, S., Schlesner, H., & Stackebrandt, E. (2006). The order *Planctomycetales*, including the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera* and the *Candidatus* genera *Brocadia*, *Kuenenia* and *Scalindua*. The Prokaryotes: A Handbook on the Biology of Bacteria, 757-793.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. Journal of bacteriology, 173(2), 697-703.

Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., & Han, S.K. (2002). Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquatic Microbial Ecology, 28(2), 141-155.