

# **Microbial Quality of Rainwater Harvested from Rooftops, for Domestic use and Homestead Food Gardens**

By

**Lizyben Chidamba**

Thesis submitted in partial fulfilment of the requirements for the degree

**DOCTOR OF PHILOSOPHY**

**(PLANT PATHOLOGY)**

In the

Department of Microbiology and Plant Pathology

Faculty of Natural and Agricultural Sciences

University of Pretoria

South Africa

**January 2015**

Supervisor: Prof. Lise Korsten

## DECLARATION

I, the under signed, declare that the thesis which I hereby submit for the degree Doctor of Philosophy in Plant Pathology at the University of Pretoria is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

.....

Lizyben Chidamba

## ACKNOWLEDGEMENTS

I sincerely wish to express my deepest gratitude and appreciation to my supervisor Prof Lise Korsten and Dr Erika Du Plessis whose support and continued guidance throughout this work made this study possible.

I am very grateful to all the staff members and students at the Plant Pathology laboratories in the Department of Microbiology and Plant Pathology, University of Pretoria for encouragement, support and assistance. In particular, I am indebted to Ms Amelita Lombard and Ms Daleen Muller for all their patience and time.

I would like to thank my wife and sons for allowing me to rob them of their family time with me so that this work could be accomplished.

I would like to thank the Water Research Commission (WRC) for funding the research project “Evaluation of the risks associated with the use of rainwater harvested from rooftops, for domestic use and, homestead food gardens; and groundwater for domestic use and livestock watering”(WRC Project No 2175/4) and the University of Pretoria for providing me with financial assistance.

## **DEDICATION**

To my loving wife Mrs Charity Chidamba and our sons Anotida and Anenyasha Chidamba.

# MICROBIAL QUALITY OF RAINWATER HARVESTED FROM ROOFTOPS, FOR DOMESTIC USE AND HOMESTEAD FOOD GARDENS

By

Lizyben Chidamba

Promoter: Prof Lise Korsten

Department: Microbiology and Plant Pathology

Faculty: Natural and Agricultural Sciences

Degree: PhD (Plant Pathology)

## RESUME

A significant number of rural households in South Africa do not have access to piped water and depend on traditional water sources including river, stream, well or pond which have potential health risks. Rainwater harvesting, a technique used to collect and store rainwater that runs off from roofs can help mitigate potable water challenges. Roof harvested rainwater (RHRW) is generally considered to be safe and is mostly used without prior treatment, which exposes people to several health risks associated with faecal pathogens. Health risks evaluation based on faecal indicator bacteria including *Escherichia coli*, faecal coliforms and enterococci revealed significant levels of contamination in RHRW. Enterococci were more prevalent in RHRW than any other faecal indicator bacteria suggesting their suitability as faecal indicator bacteria in this water system. Evaluation of *Enterococcus* spp. prevalence in pigeon faecal samples (364 isolates) showed the four species; *E. faecalis* (20.5%), *E. mundtii* (20.51%), *E. faecium* (23.1%) and *E. caseliflavus* (17.3%) to be the most prevalent, whereas *E. caseliflavus* (34.6%) and *E. mundtii* (33.2%) were more prevalent in RHRW. The use of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) for the characterisation of *E. coli* environmental isolates was evaluated as a potential tool for bacterial source tracking. Support vector machine (SVM) generated classification algorithms were able to distinguish between thirty-one different strain groups at overall accuracies between 94% and 98%. Antibiotic resistance testing among *E. coli* isolated from RHRW (109 isolates) and pigeon faecal samples (130 isolates) showed ampicillin (22.7.9%), gentamicin (23.6%), amikacin (24%), tetracycline (17.4) and amoxicillin (16.9%) to be the most frequently encountered forms of resistance. However, a relatively higher proportion of

isolates from pigeon faeces were antibiotic resistant than those from RHRW. Pyrosequencing analysis of 16S rDNA from RHRW showed the phylum *Proteobacteria* (80.5%) and the classes *Betaproteobacteria* (50.35%), *Alphaproteobacteria* (16.19%), *Verrucomicrobiae* (6.59%), *Planctomycetacia* (5.69%) and *Sphingobacteria* (3%) to be dominant. Pathogenic signatures were detected for *Acinetobacter*, *Pseudomonas*, *Clostridia*, *Chromobacterium*, *Yersinia* and *Serratia*, with *Legionella* signatures being the most prevalent in the samples. The data gathered here will allow for the implementation of a more detailed and accurate health risk assessment for the safe use of RHRW in South African rural communities.

## TABLE OF CONTENTS

DECLARATION .....	ii
ACKNOWLEDGEMENTS .....	iii
DEDICATION .....	iv
RESUME.....	v
LIST OF FIGURES .....	xii
LIST OF TABLES .....	xiv
TABLE OF ABBREVIATIONS .....	xv
CHAPTER ONE.....	1
1        GENERAL INTRODUCTION .....	1
References .....	7
CHAPTER TWO.....	15
2        REVIEW OF THE STATE OF RAINWATER HARVESTING IN SOUTH AFRICAN RURAL COMMUNITIES .....	15
Abstract.....	15
2.1      Introduction.....	16
2.2      Rainwater harvesting.....	17
2.3      Rainwater harvesting systems .....	18
2.3.1 Catchment surface .....	19
2.3.2 Gutters .....	20
2.3.3 First flush diverters.....	21
2.3.4 Storage tanks .....	22
2.4      The acceptability of rainwater harvesting .....	23
2.5      Legal aspects of rainwater harvesting .....	24
2.6      The potential impact of rainwater harvesting on water related diseases.....	24
2.7      Roof harvested rainwater quality .....	25
2.8      Conclusion and recommendations .....	27
2.9      References .....	29
CHAPTER THREE .....	35
3        STRATEGY TOWARDS THE DEVELOPMENT OF RAINWATER HARVESTING RISK ASSESSMENT AND QUALITY GUIDELINES.....	35
Abstract .....	35
3.1      Introduction.....	36

3.2	The need for risk assessment criteria and guidance information .....	37
3.3	Microbial risk assessment of roof harvested rainwater.....	38
3.4	Systems approach: identifying potential hazards and health risks.....	39
3.4.1	Microbial contaminants .....	40
3.4.2	Zoonosis .....	41
3.5	Quantitative microbial risk assessment.....	42
3.5.1	Detection of pathogens in water .....	43
3.5.2	Biofilm formation in rainwater storage tanks.....	44
3.6	Discussion .....	45
3.7	Conclusion and Recommendations .....	49
3.8	References .....	50
CHAPTER FOUR .....		60
4	A SCOPING STUDY ON PREVALENCE OF <i>ESCHERICHIA COLI</i> AND <i>ENTEROCOCCUS</i> SPECIES IN HARVESTED RAINWATER STORED IN TANKS .....	60
Abstract.....		60
4.1	Introduction.....	61
4.2	Methodology .....	64
4.2.1	Sampling sites description and sample collection .....	64
4.2.2	Microbiological analysis of water samples .....	66
4.2.3	Faecal sample collection.....	66
4.2.4	Recovery of isolates and presumptive identification.....	67
4.2.5	Matrix-assisted laser desorption ionisation time of flight mass spectroscopy identification and characterisation of bacterial isolates .....	67
4.2.6	Statistical Analysis .....	68
4.3	Results.....	68
4.3.1	Samples from Gauteng Province (Johannesburg and Pretoria).....	68
4.3.2	Prevalence of <i>Escherichia coli</i> , enterococci and <i>Pseudomonas</i> spp.in roof harvested rainwater samples from Luthengele village, Eastern Cape.....	70
4.3.3	Prevalence of enterococci in roof harvested rainwater and bird faecal samples .....	71
4.4	Discussion .....	72
4.4.1	Factors affecting rainwater quality .....	73
4.4.2	Rainwater quality from different technological and environmental settings .....	73
4.4.3	Roof harvested rainwater quality.....	75



4.4.4	Indicator bacteria in roof harvested rainwater .....	77
4.4.5	Enterococci in roof harvested rainwater .....	78
4.5	Conclusions and recommendations .....	80
4.6	References .....	81
CHAPTER FIVE .....		88
5	MALDI-TOF-MS CHARACTERISATION OF <i>ESCHERICHIA COLI</i> ISOLATED FROM ROOF HARVESTED RAINWATER .....	88
	Abstract .....	88
5.1	Introduction .....	89
5.2	Methodology .....	92
5.2.1	The Matrix-assisted laser desorption ionisation time of flight mass spectroscopy and characterisation of bacterial isolates were described in Chapter 4 .....	93
5.2.2	Polymerase chain reaction for detection of UidA gene in <i>Escherichia coli</i> .....	93
5.2.3	Data evaluation .....	93
5.3	Results .....	94
5.3.1	Optimisation of experimental parameters .....	94
5.3.2	Reproducibility of the MALDI-TOF-MS approach .....	95
5.3.3	Study work flow .....	96
5.3.4	MALDI-TOF-MS hierarchical cluster analysis of <i>Escherichia coli</i> isolates .....	96
5.3.5	Strain groups similarity evaluation .....	98
5.3.6	Machine learning .....	100
5.3.7	Classification of <i>Escherichia coli</i> strains by their sources .....	101
5.3.8	Principal component analysis .....	102
5.4	Discussion .....	107
5.4.1	Principal component analysis .....	110
5.4.2	Support vector machine .....	112
5.4.3	Comparison of MALDI-TOF-MS analysis and REP PCR analysis .....	113
5.5	Conclusion .....	114
5.6	References .....	115
CHAPTER SIX .....		121
6	ANTIBIOTIC RESISTANCE IN <i>ESCHERICHIA COLI</i> ISOLATES FROM ROOF HARVESTED RAINWATER (RHRW) TANKS .....	121
	Abstract .....	121

6.1	Introduction.....	122
6.2	Methodology .....	126
6.2.1	Area of the study and collection of samples.....	126
6.2.2	Sample collection and isolation of presumptive <i>Escherichia coli</i> strains .....	127
6.2.3	Faecal sample collection.....	127
6.2.4	Recovery.....	128
6.2.5	MALDI-TOF-MS identification and characterisation of bacterial isolates .....	128
6.2.6	Polymerase chain reaction for detection of UidA gene in <i>Escherichia coli</i> .....	129
6.2.7	Antimicrobial susceptibility testing.....	129
6.2.8	Statistical analysis .....	130
6.3	Results.....	131
6.3.1	Antibiotic susceptibility.....	131
6.3.2	Non-parametric test .....	133
6.3.3	Correlation test .....	134
6.3.4	Antibiotic resistance phenotypes .....	135
6.3.5	Cluster analysis.....	138
6.4	Discussion .....	139
6.4.1	Antibiotic resistance profiles of samples.....	142
6.4.2	Multiple antibiotic resistance patterns.....	144
6.4.3	Cluster analysis of antibiotic resistance profiles .....	146
6.5	Conclusion and recommendations for further studies.....	147
6.6	References .....	148
CHAPTER SEVEN .....		159
7	PYROSEQUENCING ANALYSIS OF ROOF HARVESTED RAINWATER AND RIVER WATER USED FOR DOMESTIC PURPOSES IN LUTHENGELE VILLAGE IN THE EASTERN CAPE PROVINCE IN SOUTH AFRICA.....	159
Abstract.....		159
7.1	Introduction.....	160
7.2	Methodology .....	163
7.2.2	DNA extraction and Pyrosequencing .....	163
7.2.3	Pyrosequencing data processing and analysis .....	164
7.2.4	Pathogen identification.....	164
7.3	Results.....	165

7.3.1	Sequencing depth.....	165
7.3.2	Alpha diversity .....	165
7.3.3	Phylum level diversity .....	167
7.3.4	Class level diversity.....	168
7.3.5	Diversity at order level .....	169
7.3.6	Diversity at family level .....	170
7.4	Cluster analysis and genus level diversity .....	171
7.4.1	Cluster analysis.....	172
7.4.2	Genus level diversity .....	173
7.4.3	UniFrac analysis .....	176
7.4.4	Detection of pathogenic signatures .....	176
7.5	Discussion .....	179
7.5.1	Microbial diversity .....	179
7.5.2	Environmental influence on microbial diversity .....	182
7.5.3	Bacterial diversity.....	183
7.5.4	Detection of pathogenic signatures .....	184
7.6	Conclusion and recommendations .....	187
7.7	References .....	188
CHAPTER EIGHT .....		197
8	GENERAL DISCUSSION .....	197
References .....		203

## LIST OF FIGURES

Figure 4.1: Prevalence of bacterial species in tank and kitchen water samples .....	70
Figure 4.2: Prevalence of bacterial species in ground harvested rainwater used for irrigation and river and spring water used for potable purposes. ....	71
Figure 4.3: Distribution of enterococci isolated from pigeon faeces and roof harvested rainwater. ....	72
Figure 5.1: Matrix-assisted laser desorption ionisation time of flight mass spectroscopy spectra reproducibility .....	95
Figure 5.2: Two-dimensional unsupervised hierarchical cluster analysis of <i>Escherichia coli</i> isolates. ....	97
Figure 5.3: Mean spectrum dendrogram of the thirty one straingroups identified from two- dimensional unsupervised hierarchical cluster analysis of <i>Escherichia coli</i> isolates.....	99
Figure 5.4: Principal component analysis 3D view and explained variance. ....	104
Figure 5.5: Matrix-assisted laser desorption ionisation time of flight mass spectral stack view of selected peaks derived from the original profile according to principal component analysis.. ....	105
Figure 5.6: The dimensional image from principal component analysis generated with minimised profile data, showings the distinction between thirty-one <i>Escherichia coli</i> strain groups.. ....	107
Figure 6.1: The variability in antibiotic resistance/susceptibility patterns for where significant differences observed between the different sources .....	134
Figure 6.2: Abundance percentages of the observed phenotypes per groups of antibiotics by their sources.....	137
Figure 6.3: Dendrogram showing the relationship between <i>Escherichia coli</i> isolated from pigeon faeces and roof harvested rainwater samples obtained from sites in Johannesburg and Pretoria.....	138
Figure 7.1: Rarefication curves for species richness and diversity at genetic distances. ....	166
Figure 7.2: Average relative abundance of the reads classified into the predominant microbial phyla by individual samples. ....	168
Figure 7.3: Average relative abundance of the reads classified into the predominant microbial classes of individual samples.....	169
Figure 7.4: Average relative abundance of the reads classified into the predominant microbial classes of individual samples.....	170

Figure 7.5: Relative abundance of sequences classified into respective bacteria families of the individual samples.....	171
Figure 7.6: Clustering of samples. Bray-Curis similarity index was calculated using the abundance of operational taxonomic units and hierarchical clustering was calculated using the single linkage and Euclidean algorithms .....	172
Figure 7.7: Schematic drawing showing the detected bacterial genera found in the three different groups of water samples from Luthegele village.....	174
Figure 7.8 Bacterial communities of river water, kitchen water and tank water.....	175
Figure 7.9: Principal coordinate analysis of unweighted UniFrac distances.....	176
Figure 7.10: Evolutionary relationships of taxa .....	178

## LIST OF TABLES

Table 2.1: List of challenges and opportunities in roof rainwater harvesting systems .....	27
Table 5.1: Identification results of a support vector machine model on the binary peak data of the thirty-one unique strain groups. . . . .	100
Table 5.2: Frequency of occurrence of <i>Escherichia coli</i> strains from different sources in relation to homology/similarity groups observed from the dendrogram. ....	102
Table 5.3: Data for the eleven peaks obtained from the loading plots of the principal component analysis data.....	106
Table 6.1: Details of antibiotics that were used in the study to test for antibiotic resistance.	130
Table 6.2: Antibiotic resistance profiles of <i>Escherichia coli</i> isolated from roof harvested rainwater originating from Pretoria and Johannesburg, and bird faecal samples in Gauteng Province of South Africa.....	132
Table 6.3: Kruskal Wallis test of differences between pigeon and roof harvested rainwater isolate sources.....	133
Table 1.4: Pearson’s product moment correlation coefficient between different sites, areas and sources of <i>Escherichia coli</i> antibiotic resistance profiles.....	135
Table 6.5: Antibiotic resistance phenotypes of <i>Escherichia coli</i> isolated from pigeon faeces and roof harvested rainwater obtained from rainwater harvesting tanks.....	136
Table 7.1: Alpha diversity parameters of bacterial communities from river and roof harvested rainwater samples. ....	167
Table 7.2: Probable species of the detected genera known to contain pathogenic species ....	177

## TABLE OF ABBREVIATIONS

AAAE	African Association of Agricultural Economists
AEASA	Agricultural Economics Association of South Africa
AGT	Above-ground tanks
ANN	Artificial neural network
AR	Antibiotic resistance
ARA	Antibiotic resistance analysis
ASM	American Society for Microbiology
BST	Bacterial source tracking
CLSI	Clinical and Laboratory Standards Institute
DGGE	Denaturing gradient gel electrophoresis
DRWH	Domestic rainwater harvesting
DWAF	Department of Water Affairs and Forestry
EPA	Environmental Protection Agency
FIB	Faecal indicator bacteria
FISH	Fluorescence in situ hybridisation
GA	Genetic algorithm
GRWH	Ground surface rainwater harvesting
HIV	Human immune deficiency virus
MALDI	Matrix-assisted laser desorption ionisation
MAR	Multiple antibiotic resistance
MDG	Millennium Development Goals
MRA	Microbial risk assessment
MSP	Municipal services project
MST	Microbial source tracking
NRDC	Natural Resources Defence Council
OTU	Operational taxonomic units
PCA	Principle component analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
POU	Point of use
PSJ	Port St Johns
QMRA	Quantitative microbial risk assessment

RDI	Reference database isolates
RRWH	Roof rainwater harvesting
RWH	Rainwater harvesting
SD	Similarity distance
SEI	Stockholm Environment Institute
SSU	Small sub-unit
SVM	Support vector machine
UPGMA	Unweighted pair groups' method analysis
USDA	United States Department of Agriculture
UV	Ultra violet
VBNC	Viable but non-culturable
WISA	Water Institute of Southern Africa
WPGMA	Weighted pair groups method analysis
WRC	Water Research Commission



## CHAPTER ONE

### GENERAL INTRODUCTION

Globally, water is a scarce resource and an estimated 41 per cent of the world's population (or 2.3 billion people), live under 'water stress', while 1.1 billion people live without access to safe drinking water (FitzMaurice, 2007). Access to clean water is worst in developing countries with at least one third of the population living without access to safe drinking water and about 1.87 million children die annually due to diarrhoea (Boschi-Pinto et al., 2008).

In South Africa, the demand for water is in excess of natural water availability in several river basins, making it a water scarce country (Oberholster and Ashton, 2008; van Vuuren, 2008; Roux et al., 2010). The country has different climates with variable rainfall patterns and high evaporation rates (Kahinda et al., 2009; Stockholm Environment Institute (SEI), 2009; Everson et al., 2011). Much of South Africa has predominantly hard rock geology, which limits groundwater availability. Hence, surface water is the most significant resource (van der Merwe-Botha, 2009). However, surface water has become very contaminated due to mining, industrial and agricultural activities and informal settlements next to riverbanks. Hence there is a need to evaluate other alternative sources of clean freshwater (Roux et al., 2010).

Social and demographic factors also contribute to water scarcity. One of these factors, for instance, is the distribution of significant settlements and industry that is determined by mineral deposits rather than water resources. In areas where groundwater is available, it is frequently over-exploited (van der Merwe-Botha, 2009). Worse still, population and economic growth in South Africa has resulted in increased demand on freshwater resources, including groundwater, man-made lakes and rivers (Ochse, 2007; Oberholster and Ashton, 2008; Kahinda et al., 2010; Viljoen et al., 2012). The quality of water from these resources has declined significantly due to pollution from urbanisation, agriculture, industries, mining and power generation (Department of Water Affairs and Forestry, 2002; Ochse, 2007; Roux et al., 2010). Given the current patterns of water use and discharge, anticipated future population growth rates and expected socio-economic development trends, it is most likely

that the available water resources will not be sufficient for future needs (Dalvie et al., 2003; Oberholster and Ashton, 2008; van der Merwe-Botha, 2009; Roux et al., 2010). It has been forecasted that freshwater resources in South Africa will be depleted and unable to meet the needs of industry and the people by the year 2030 (Postel, 2000; Turton, 2003).

In as much, as the South African government has made great strides in the provision of clean domestic water, many poor and vulnerable South African inhabitants either have access to insufficient water or the available water is not of suitable quality for drinking or personal hygiene (Statistics South Africa, 2010). The problems of inadequate supplies and insufficient treatment encourage searching for decentralised alternative approaches to access clean domestic and agricultural water, keeping in mind the technical and financial limitations of the poor living in under-developed areas (Alcock and Verste, 1987; Kahinda et al., 2007, 2010; Bulcock and Schulze, 2011; Kahinda and Taigbenu, 2011).

Domestic rainwater harvesting (DRWH) describes the small-scale concentration, collection, storage, and use of rainwater runoff for production purposes. Roof rainwater harvesting (RRWH) is one of the broad categories of DRWH where water is collected from roofs, and stored in underground tanks (UGTs) or above-ground tanks (AGTs) and used for domestic purposes, including small scale production activities such as garden watering (Kahinda et al., 2007, 2010). Roof rainwater harvesting is one of the most appropriate alternative sources of potable and non-potable water supplies at household or community level as the world faces decreasing water sources and increasing energy crisis (Amin and Han, 2009).

Roof rainwater harvesting is an ancient technique for collecting and storing rainwater during the rainy season for later use. The practice of rainwater harvesting is verified by archaeological evidence which date back as far as 4000 years ago (Richards, 1989). Rainwater Harvesting (RWH) is not a conservation technique but rather a new water supply (Critchley and Siegert, 1991). Now that people are faced with decreasing water sources and increasing energy crisis, roof harvested rainwater RRWH may represent an alternative water source for supplying freshwater at household or community level (Houston and Still, 2002; Baiphethi et al., 2010; Stimie et al., 2010; Bulcock and Schulze, 2011; Botha et al., 2012;

Viljoen et al., 2012). Roof rainwater harvesting has received increased attention worldwide as an alternative source of potable and non-potable water supplies. Potential applications of RRWH do exist for roof catchments in areas where a centralised water supply and distribution system are not adequate, and such applications are increasing (Han and Mun, 2007, 2011).

Many regions around the world are adopting RRWH to overcome the increasing demand of water in addition to climate change adaptation (Jackson et al., 2001). Roof rainwater harvesting offers benefits such as promoting self-sufficiency, encouraging water and energy conservation and results in a permanent decrease in water demand from the main water supply (Retamal and Turner, 2010). Roof harvested rainwater has been used as a non-potable water source in many places around the world (Critchley and Siegert, 1991; Retamal and Turner, 2010). The use of rainwater in domestic hot water systems can produce optimum environmental and economic results for urban water cycle management. In Palestine, RRWH systems have been shown to reduce the annual environmental impact of the in-house water usage by about 40% (Spinks et al., 2006a). Roof harvested rainwater use for supplying drinking water in urban areas has a long history in semi-arid areas (Boers, 1994). A recent case study in Sudan has shown RHRW to provide an additional source of drinking water in a changing physical environment associated with urban population growth (Ibrahim, 2009).

Harvesting rainwater from roofs is an ecological friendly alternative approach to addressing the country's critical water shortage needs. Water collected in this manner can address domestic water shortage and provide irrigation water for domestic food gardens (Kahinda et al., 2007; Kahinda and Taigbenu, 2011). Prior to promoting RRWH, it is essential to determine potential microbiological risks that can be associated with such water collection systems. Water collected in this manner is commonly stored in large plastic, metal or ferrocement containers. The ability of microorganisms to proliferate in such water storage systems has been well documented (Simmons et al., 2008; Coombes, 2002). The quality of harvested and stored rainwater is however, not well known. In general, dust, bird droppings, chemical leachates from the roofing material, adhesives, and coatings, etc., may be leached and washed down from the roofs during rain events into storage containers, posing a potential health risk for the consumer (Mendez et al., 2011; Lee et al., 2012). Water quality may thus be

compromised by the water collection approach. In addition, biofilms may develop in the storage unit further compromising the water quality (Evans et al., 2006b; Kim and Han, 2011; Domènech et al., 2012). This is of particular importance since it is known that water-borne pathogens may survive, proliferate and shed from biofilms into water thereby contributing to the contamination risk (Donlan and Costerton, 2002; U. S. Environmental Protection Agency, 2002).

Faecal deposition on the catchment surface is the main source of microbiological contamination. Mild to severe gastroenteritis may be caused by numerous infectious agents, including enteric bacteria, viruses, and protozoa (Schets et al., 2010). Rainwater collected from a well maintained catchment (clean roof surfaces and gutters) may provide water of acceptable quality (Cunliffe, 1998); however some studies have shown that the microbial quality of rainwater stored in tanks can be very poor (Ahmed et al., 2011). Overall reported levels of bacterial indicator organisms in rainwater tanks demonstrate the variability and susceptibility of collected rainwater to faecal contamination (Evans et al., 2006a).

Several studies have isolated *Salmonella*, *Clostridium perfringens*, *Aeromonas*, *Vibrio parahaemolyticus*, *Campylobacter*, *Giardia* and *Legionella*-like organisms from rainwater tank samples (Ahmed et al., 2011). While these organisms do not affect the aesthetic properties of the water, they may cause serious illness especially among immune-compromised individuals including the elderly and children (Ahmed et al., 2012).

Major outbreaks of water-borne disease attributable to rainwater are infrequently reported, probably because most tanks serve only single households. Several studies have however, reported illness linked with the consumption of rainwater (Heyworth, 2006). An outbreak of *Salmonella* Saintpaul attributed to contamination of rainwater was reported at a construction site in Queensland, Australia (Taylor et al., 2000). In Trinidad, *Salmonella* arechevalata was reported to have infected sixty-three people at a campsite and the organism was isolated from the outlet tap of the rainwater tank (Dean and Hunter, 2012). Outbreaks attributed to *Campylobacter* (Merritt et al., 1999) and *Legionella* (Simmons et al., 2008) have also been reported. In these case control studies, a strong association could be observed between illness

and consumption of harvested rainwater. Generally, outbreaks attributed to *Campylobacter* and *Salmonella* appear to have been caused by faecal contamination of the roof surface by birds or animals (Sinclair et al., 2005).

An estimated one third of the 51 million deaths worldwide annually result from infectious and parasitic diseases (Mathers et al., 2008). It has further been estimated that 50 000 people die daily worldwide as a result of water related diseases (Prüss et al., 2002). Foodborne diseases occur as a result of food contamination that may occur at any point along the food chain from the farm through food processing to the end user (Doyle and Buchanan, 2013). However, the majority of foodborne outbreaks have been traced to contamination on the farm and at times linked to the use of contaminated water (Westcot, 1997).

Having access to safe, potable water is a basic human right. However, water quality is seriously compromised by ineffective water management systems and lack of adequate control of industrial and other environmental pollutants (Yadav et al., 2007; Farahbakhsh et al., 2009; Garrison et al., 2011). Providing safe quality water is one of the basic responsibilities of government and is important in terms of agricultural activities and food safety (van der Merwe-Botha, 2009; Gemmell and Schmidt, 2010; Viljoen et al., 2012). By polluting the environment, the microbial risk to the consumer increases and has long-term negative effects on the population, food security and sustainability (van der Merwe-Botha, 2009). By understanding the risks associated with the use of RHRW, improved usage of this valuable resources can be made through intervention strategies, guidelines and regulations. Hence, basic public health issues can be managed and exposure contamination prevented. Promoting the concept of domestic food gardens can further contribute to national healthy life style and general well-being.

Food security is a critical challenge in a number of countries, including in South Africa. In order to improve the health of the nation, the consumption of fresh produce has been greatly encouraged (Food and Agriculture Organisation/World Health Organisation (FAO/WHO), 2004; Ruel et al., 2005; United States Department of Agriculture (USDA), 2008). However the increased consumption of raw fruit and vegetables been linked to disease outbreaks which

have resulted in significant losses for industry and governments (Buck and Pathology, 2003; Food and Agriculture Organisation/World Health Organisation, 2004; Valadez et al., 2012; U. S. Department of Health and Human Services: Centers for Disease Control and Prevention (USDA/CDCP), 2013). The recent *E. coli* outbreak linked to water/food in the European Union has highlighted the economic importance of having effective regulation and communication systems (Buchholz et al., 2011; King et al., 2012).

In South African rural communities vegetable gardens are promoted as a way to be self-sustainable (Kahinda et al., 2007, 2010; Denison and Wotshela, 2009; Baiphethi et al., 2010; Everson et al., 2011). Given the potential risk associated with the use of contaminated water, small scale vegetable growers and farmers stand the risk of consuming or selling contaminated produce, thereby severely affecting the health of the area population (Department of Water Affairs and Forestry 2002; Paulse et al., 2009). Although, the collection of water from roofs or surface runoff for domestic use has been encouraged, little is known about the quality of such water (Alcock and Verste, 1987; Houston and Still, 2002; Kahinda et al., 2007; Viljoen et al., 2012).

This study evaluates current literature on the implementation of rainwater harvesting practices in South Africa including risk assessment, microbial safety practices and the microbial quality of harvested rainwater used by rural communities in South Africa. We aimed to evaluate the prevalence of faecal indicator bacteria *E. coli* and enterococci in RHRW and their suitability as sanitary indicator bacteria in RHRW. We also evaluated the implication and prevalence of antibiotic resistant *E. coli* isolated from RHRW water and bird faecal samples as the most likely sources of contamination. The potential application of matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI TOF MS) based bacterial fingerprinting to distinguish between environmental strains *E. coli* was evaluated as high throughput technique in bacterial source tracking. We also aimed to assess the applicability of pyrosequencing analysis technique to characterise microbial community structures in RHRW and to detect potential pathogens.

It was anticipated that this study will provide data that will inform society of the potential

risks associated with the use of harvested rainwater. This will promote the safe use and appropriate applications of rainwater harvesting technology. The information should enable relevant government departments to develop effective intervention strategies and establish appropriate guidelines and regulations. This study was undertaken with special attention to rural communities without access to clean piped water.

## References

- Ahmed, W., J. P. S. Sidhu, S. Toze, and L. Hodgers. 2012. Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland. *Applied and Environmental Microbiology* 5: 219–226.
- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Alcock, P. G., and E. Verste. 1987. Investigation into unconventional sources of water for a peri-urban/rural district of KwaZulu. *South African Journal of Science* 83: 348–352.
- Amin, M. T., and M. Y. Han. 2009. Roof harvested rainwater for potable purposes: application of solar collector disinfection (SOCO-DIS). *Water Research* 43: 5225–35.
- Baiphethi, M. N., M. F. Viljoen, G. Kundhlande, and N. G. Ralehlolo. 2010. Enhancing food and livelihood security in the context of the food and financial crisis: challenges and opportunities for small scale rainwater harvesting and conservation. In the Joint AAAE Third Conference/AEASA 48th Conference, 19-23 September, Cape Town, South Africa (No. 96640). African Association of Agricultural Economists (AAAE) and Agricultural Economics Association of South Africa (AEASA).
- Berger, C. N., S. V. Sodha, R. K. Shaw, P. M. Griffin, D. Pink, P. Hand, and G. Frankel. 2010. Minireview: Fresh fruits and vegetables as vehicles for the transmission of human pathogens. *Environmental Microbiology* 12: 2385–2397.
- Boers, T. M. 1994. Rainwater harvesting in arid and semi-arid zones. International Institute for Land Reclamation and Improvement. Wageningen, The Netherlands.
- Boschi-Pinto, C., L. Velebit, and K. Shibuya. 2008. Estimating child mortality due to diarrhoea in developing countries. *Bulletin of the World Health Organization* 86: 710–717.

- Botha, J. J., J. J. Anderson, L. F. Joseph, R. M. Snetler, N. Monde, F. Lategan, N. N. Nhlabatsi, M. S. Lesoli, and S. Dube. 2012. Manual sustainable techniques and practices for water harvesting and conservation and their resource-poor agricultural production. WRC Report No. TT 542/12. Water Research Commission (WRC). Pretoria.
- Buchholz, U., H. Bernard, D. Werber, M. M. Böhmer, C. Remschmidt, H. Wilking, Y. Deleré, M. Der Heiden, C. Adlhoch, J. Dreesman, J. Ehlers, S. Ethelberg, M. Faber, C. Frank, G. Fricke, M. Greiner, M. Höhle, S. Ivarsson, M. Sc, U. Jark, M. Kirchner, J. Koch, G. Krause, P. Lubert, B. Rosner, and K. Stark. 2011. German Outbreak of *Escherichia coli* O104:H4 Associated with Sprouts. *The New England Journal of Medicine* 365: 1763–1770.
- Buck, J. W., R. R. Walcott and L. R. Beuchat, 2003. Recent trends in microbiological safety of fruits and vegetables. *Plant Health Progress*. doi: 10.1094/PHP-2003-0121-01-RV. <https://www.apsnet.org/publications/apsnetfeatures/Pages/microsafety.aspx>
- Bulcock, L. M., and R. E. Schulze. 2011. Climate change and rainwater water harvesting in South Africa: A Case Study. Paper presented at the 16<sup>th</sup> Paper SANCIAHS national hydrology symposium, 1-3 October 2012, Pretoria, South Africa.
- Coombes, P. J. 2002. Rainwater tanks revisited: new opportunities for urban water cycle management. Phd Thesis, University of Newcastle, Callaghan, Australia.
- Critchley, W., and K. Siegert. 1991. Water harvesting. Food and Agriculture Organisation of the United Nations, Rome, Italy: 133.
- Cunliffe, D. A. 1998. Guidance on the use of rainwater tanks. National Environmental Health Forum Monographs. Water Series 5210: 1–29.
- Dalvie, M. A., E. Cairncross, A. Solomon, and L. London. 2003. Contamination of rural surface and ground water by endosulfan in farming areas of the Western Cape, South Africa. *BioMed Central. Environmental Health* 2: 1-15.
- Dean, J., and P. R. Hunter. 2012. Risk of gastrointestinal illness associated with the consumption of rainwater: A systematic review. *Environmental Science Technology* 46: 2501–7.



- Denison, J., and L. Wotshela. 2009. Indigenous water harvesting and conservation practices : Historical context, cases and implications. WRC Report No. TT 392/09. Water Research Commission (WRC). Pretoria.
- Department of Water Affairs and Forestry. 2002. National microbial monitoring programme for surface water: Implementation Manual. Pretoria, South Africa.
- Domènech, L., H. Heijnen, and D. Saurí. 2012. Rainwater harvesting for human consumption and livelihood improvement in rural Nepal: benefits and risks. *Water and Environment Journal* 26: 465–472.
- Donlan, R. M., and J. W. Costerton. 2002. Biofilms: Survival mechanisms of clinically relevant microorganisms. *Clinical Microbiology Reviews* 15: 167–193.
- Doyle, M. P., and Buchanan, R. L. (eds) 2013. *Food microbiology: Fundamentals and frontiers* (4<sup>th</sup> Edition). ASM Press. Washington DC.
- Evans, C. A., P. J. Coombes, and R. H. Dunstan. 2006a. Wind, rain and bacteria: The effect of weather on the microbial composition of roof harvested rainwater. *Water Research* 40: 37–44.
- Evans, C. A., P. J. Coombes, R. H. Dunstan, T. Harrison, A. Martin, and J. N. Harris. 2006b. Coliforms, biofilms, microbial diversity and the quality of roof harvested rainwater. *Water Research* 40: 34–44.
- Everson, C., T. M. Everson, A. T. Modi, D. Csiwila, M. Fanadzo, V. Naiken, R. M. B. Auerbach, M. Moodley, S. M. Mtshali, and R. Dladla. 2011. Sustainable techniques and practices for water harvesting and conservation and their effective application in resource- poor agricultural production participatory adaptive research through participatory adaptive research. WRC REPORT NO. 1465/1/11. Water Research Commission, Pretoria.
- Farahbakhsh, K., C. Despins, and C. Leidl. 2009. Developing capacity for large-scale rainwater harvesting in Canada. *Water Quality Research Journal of Canada* 44: 92–102.
- Fitzmaurice, M. 2007. The human right to water. *Fordham Environmental Law Review* 18: 537–573.
- Food and Agriculture Organisation/World Health Organisation. 2004. Fruit and vegetables. Report of a Joint FAO/WHO Workshop, 1–3 September 2004, Kobe, Japan: 1–3.

- Garrison, N., N. C Kloss, R. Lukes, J. Devine, and S. Beckman. 2011. Capturing rainwater from rooftops : an efficient water resource management strategy that increases supply and reduces pollution. Natural Resources Defense Council (NRDC) New York.
- Gemmell, M. E., and S. Schmidt. 2010. Potential links between irrigation water quality and microbiological quality of food in subsistence farming in KwaZulu-Natal, South Africa. In: Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology: edited by A Mendez-Vilas, Formatex Research Centre, Bodajaz, Spain. pp 1190–1195.
- Han, M. Y., and J. S. Mun. 2011. Operational data of the Star City rainwater harvesting system and its role as a climate change adaptation and a social influence. *Water Science Technology* 63: 2796.
- Han, M. Y., and J. S. Mun. 2007. Particle behaviour consideration to maximise the settling capacity of rainwater storage tanks. *Water Science and Technology* 56: 73–79.
- Heyworth, J. S., G. Glonek, E. J. Maynard, P. A. Baghurst, and J. Finlay-Jones. 2006. Consumption of untreated tank rainwater and gastroenteritis among young children in South Australia. *International Journal of Epidemiology* 35: 1051–1058.
- Houston, P. C., and D. A. Still. 2002. Rain water harvesting: A neglected rural water supply option. In Biennial Conference of the Water Institute of Southern Africa (WISA). 19 – 23 May 2002, Durban, South Africa.
- Ibrahim, M. B. 2009. Rainwater harvesting for urban areas: A success story from Gadarif City in Central Sudan. *Water Resources Management* 23: 2727–2736.
- Jackson, R. B., S. R. Carpenter, C. N. Dahm, D. M. McKnight, R. J. Naiman, S. L. Postel, and S. W. Running. 2001. Water in a changing world. *Ecological Applications* 11: 1027–1045.
- Kahinda, J. M., A. E. Taigbenu, and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 32: 1050–1057.
- Kahinda, J. M., A. E. Taigbenu, and R. J. Boroto. 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.

- Kahinda, J. M., A. E. Taigbenua, E. S. B. Lillieb, and R. J. Boroto. 2009. A GIS-based decision support system for rainwater harvesting (RHADESS). *Physics and Chemistry of the Earth Parts ABC* 34: 767–775.
- Kahinda, J. M., and A. E. Taigbenu. 2011. Rainwater harvesting in South Africa: Challenges and opportunities. *Physics and Chemistry of the Earth Parts ABC* 36: 968–976.
- Kim, M., and M. Han. 2011. Composition and distribution of bacteria in an operating rainwater harvesting tank. *Water science and technology* 63: 1524–1530.
- King, L. A., F. Nogareda, P. Mariani-kurkdjian, E. Loukiadis, G. Gault, N. Jourdan-DaSilva, E. Bingen, M. Mace, D. Thevenot, N. Ong, C. Castor, H. Noel, D. Van Cauteren, M. Charron, V. Vaillant, B. Aldabe, V. Goulet, G. Delmas, E. Couturier, Y. Le Strat, C. Combe, Y. Delmas, F. Terrier, B. Vendrely, P. Rolland, and H. de Valk. 2012. Outbreak of shiga toxin producing *Escherichia coli* O104: H4 associated with organic fenugreek sprouts, France, June 2001. *Clinical Infectious Diseases* 54: 1588–1594.
- Lee, J. Y, G. Bak and M. Han. 2012. Quality of roof harvested rainwater - comparison of different roofing materials. *Environmental Pollution* 162: 422–429.
- Mathers, C., Fat, D. M., and Boerma, J. T. 2008. The global burden of disease: 2004 update. World Health Organisation.
- Mathipa, K. S. 2008. The development of an environmental education programme for water conservation in the Steelpoort area. Msc Dissertation, University of South Africa. Pretoria.
- Mendez, C. B., J. B. Klenzendorf, B. R. Afshar, M. T. Simmons, M. E. Barrett, K. A. Kinney, and M. J. Kirisits. 2011. The effect of roofing material on the quality of harvested rainwater. *Water Research* 45: 2049–2059.
- Merritt, A., R. Miles, and J. Bates. 1999. An outbreak of *Campylobacter enteritis* on an island resort, North Queensland. Australian Government Publishing Service. *Communicable Diseases Intelligence* 23: 215–219.
- Oberholster, P. J., and P. J. Ashton. 2008. State of the Nation Report: An overview of the current status of water quality and eutrophication in South African rivers and reservoirs. Parliamentary Grant Deliverable. Pretoria, South Africa.

- Ochse, M. 2007. Seasonal rainfall influences on main pollutants in the Vaal river barrage reservoir : A temporal-spatial perspective. Msc Dissertation, University of Johannesburg. South Africa
- Paulse, A. N., V. A. Jackson, and W. Khan. 2009. Comparison of microbial contamination at various sites along the Plankenburg and Diep Rivers, Western Cape, South Africa. *WaterSA* 35: 469–478.
- Postel, S. L. 2000. Entering an era of water scarcity: the challenges ahead. *Ecological Applications* 10: 941–948.
- Prüss, A., D. Kay, L. Fewtrell, and J. Bartram. 2002. Estimating the burden of disease from water, sanitation, and hygiene at a global level. *Environmental Health Perspectives* 110: 537–542.
- Retamal, M., and A. Turner. 2010. Unpacking the energy implications of distributed water infrastructure: how are rainwater systems performing? *Water Science and Technology: Water Supply* 10: 546–553.
- Richards, T. S. 1989. Evidence of ancient rainwater concentrating structures in northern Egypt as seen on Landsat MSS imagery. *International Journal of Remote Sensing* 10: 1135–1140.
- Roux, S., S. Oelofse, and W. De Lange. 2010. Can South Africa afford to continue polluting its water resources? – With special reference to water pollution in two important catchment areas. CSIR 3rd Biennial Conference 2010. Science Real and Relevant, Council for Scientific and Industrial Research (CSIR) International Convention Centre, Pretoria.
- Ruel, M. T., N. Minot, and L. Smith. 2005. Patterns and determinants of fruit and vegetable consumption in sub-Saharan Africa: A multicountry comparison. International Research institute. World Health Organisation.
- Schets, F. M., R. Italiaander, H. H. J. L. Van Den Berg, and A. M. De RodaHusman. 2010. Rainwater harvesting: Quality assessment and utilisation in The Netherlands. *Journal of Water and Health* 8: 224–235.

- Simmons, G., S. Jury, C. Thornley, D. Harte, J. Mohiuddin, and M. Taylor. 2008. A Legionnaires' disease outbreak: A water blaster and roof-collected rainwater systems. *Water Research* 42: 1449–1458.
- Sinclair, M., K. Leder, and H. Chapman. 2005. Public health aspects of rainwater tanks in urban Australia. CRC for Water Quality and Treatment.
- Statistics South Africa (Stats SA). 2010. General household survey. Statistical Release, May 2007 Pretoria.
- Stimie, C. M., E. Kruger, M. de Lange, and C. T. Crosby. 2010. Agricultural water use in homestead gardening systems Volume 1 : Main Report. WRC Report No: TT 430/09. Water Research Commission (WRC). Pretoria.
- Stockholm Environment Institute (SEI). 2009. Rainwater harvesting: A lifeline for human well-being. Job No. DEP/1162/NA UNEP. United Nations Environment Programme.
- Taylor, R., D. Sloan, T. Cooper, B. Morton, I. Hunter, and others. 2000. A water-borne outbreak of *Salmonella* Saintpaul. *Communicable diseases intelligence* 24: 336–339.
- Turton, A. R. 2003. The political aspects of institutional developments in the water sector: South Africa and its international river basins. PhD Thesis, University of Pretoria, South Africa. .
- U. S. Department of Health and Human Services: Centres for Disease Control and Prevention (USDA/CDCP). 2013. Surveillance for food borne disease outbreaks — United States, 1998 – 2008. *Morbidity and mortality weekly report. Surveillance Summaries* 62: 1–34.
- U. S. Environmental Protection Agency. 2002. Health risks from microbial growth and biofilms in drinking water distribution systems. U. S. Environmental Protection Agency Office of Ground Water and Drinking Water Standards and Risk Management, Pennsylvania, Washington DC.
- US Department of Agriculture (USDA). 2008. Increasing fruit and vegetable consumption through the USDA nutrition assistance programs. Progress Report. US Department of Agriculture Food and Nutrition Service Office of Research, Nutrition, and Analysis., Pennsylvania, Washington DC.

- Valadez, A. M., K. R. Schneider, and M. D. Danyluk. 2012. Outbreaks of foodborne diseases associated with tomatoes: Series FSHN12-08. Retrieved April 1, 2013, from <http://edis.ifas.ufl.edu/fs192>
- Van der Merwe-Botha, M. 2009. Water Quality: A vital dimension of water security. Development Planning Division. Working Paper Series No. 14, Development Bank Of Southern Africa: Midrand.
- Van Vuuren, L. 2008. Start saving or start paying, river studies warn : water resources planning. *Water Wheel* May/June 2008 7: 14–18.
- Viljoen, M. F., G. Kundhlande, M. N. Baiphethi, P. Esterhuyse, J. J. Botha, J. Minkley, and G. A. Anderson. 2012. An assessment of the social and economic acceptability of rainwater harvesting and conservation practices in selected peri-urban and rural communities. WRC Report No. 1648/1/12. Pretoria, South Africa.
- Westcot, D. W. 1997. Quality control of wastewater for irrigated crop production (Water reports-10). Food and Agriculture Organisation of the United Nations, Rome, Italy.
- Yadav, S., R. Singhvi, and B. K. Sharma. 2007. Recharging of borewells and analysis of harvested rooftop rainwater in houses of Udaipur city. *Journal of Environmental Science Engineering* 49: 225–228.

## CHAPTER TWO

# REVIEW OF THE STATE OF RAINWATER HARVESTING IN SOUTH AFRICAN RURAL COMMUNITIES

### **Abstract**

A significant number of South African rural households still depend on traditional water sources including river, spring, pond or dam which have potential health associated risks. The task of improving access to clean water among these communities is challenging given that the installation of piped water services is prohibitively expensive. This is mainly due to the often difficult terrain and large distances between households. Consequently, the collection and storage of rainwater that runs off from rooftops can help mitigate potable water challenges. However, rainwater harvesting is still not clearly defined in government planning, policies and regulations although it is widely promoted. A number of factors affect harvested rainwater (RHRW) quality, such as chemical and microbial contamination. Furthermore RHRW is generally used directly without prior treatment, which exposes people to potential health risks. Contamination of RHRW rainwater is mainly due to dust and faecal material present on the catchment surface. The potential for microbial hazards in potable water to affect consumers are dependent on the barriers that are in place. These start with the prevention of contamination at source and at various stages thereafter, including storage and distribution. It is therefore vital that appropriate guidelines are available to ensure that the potential health risks associated with the use of RHRW are minimised, while the benefits are maximised.

**Key Words:** rural communities, rainwater harvesting, contamination, health risk

## 2.1 Introduction

The South African government is one of the signatories of the Millennium Development Goals (MDG) which require that the proportion of people without sustainable access to safe drinking water and basic sanitation be halved by 2015 (World Health Organisation, 2008). The declaration further stipulates that the first six kilolitres of water consumed monthly by poor households (with less than R1120 (= US\$ 112) income/month) be provided free of charge (MDG 7, Target 3) (World Health Organisation, 2008). However, water resources and hydrological systems in South Africa are generally under stress as a result of low rainfall and high evaporation rates (Mukheibir, 2008). This is further worsened by the ever-growing demand for water, emanating from population growth and economic development policies pertaining to water resources (Webster, 1999).

Although it is politically plausible that everyone will receive piped water services, homesteads in most rural areas are scattered and the terrain maybe steep, making piped water access to be prohibitively expensive (Cothren, 2013). Consequently, most rural areas are without piped water and where it is present, the supply is usually erratic. This result in households resorting to traditional river, stream, well or pond water sources (Nevondo and Cloete, 1999). The use of these water sources for drinking purposes has potential health risks since they may be contaminated with human or animal faeces (Lehloesa and Muyima, 2000). Consequently, a significant proportion of residents in rural communities in South Africa are exposed to water-borne diseases including campylobacteriosis, shigellosis, salmonellosis, cholera, yersiniosis and a variety of other bacterial, as well as fungal, viral and parasitic infections (Obi and Bessong, 2011). These diseases can cause crippling, devastating and debilitating effects in poor communities and further exacerbate the already strained health system and facilities in the country (Obi et al., 2002).

The need for access to clean domestic water and the impact of not providing it were revealed by the cholera epidemic in 2000–2001 (Mugero and Hoque, 2001). A total of 265 deaths and 117,147 infections were reported in South Africa for that period (Mugero and Hoque, 2001; Hemson et al., 2006). A review of the state of water provision in rural areas and informal settlements where the epidemic was worst, indicated that the then government policies of cost



recovery resulted in a shift back to the use of traditional, untreated water sources (Cottle and Deedat, 2002). Although the government's response to the cholera epidemic has led to improved provision of safe water and sanitation to the poor, events in the period since the outbreak have highlighted the need for effective service delivery. The typhoid outbreak in Delmas, South Africa in 2005, where five deaths and 596 cases of the disease, accompanied by 3346 cases of diarrhoea were reported demonstrated the continued vulnerability of poor people in urban and rural settings (Sidley, 2005).

Although services have been improved in many areas, underlying poverty remains a reality for most of the historically disadvantaged people (Kahinda et al., 2007). It is therefore imperative to provide clean water sources, regularly monitor microbial quality of the water, identify potential health risks and provide the impetus for sustained government intervention. It is important to note that the supply of water in rural communities is one of the great challenges of sustainable development because it impinges on achieving the objectives of improving health, income, living conditions and ensuring equitable and sustainable use of natural resources and a better life for all in South Africa (Kahinda and Taigbenu, 2011; Viljoen et al., 2012).

## **2.2 Rainwater harvesting**

The availability of a water source that can be controlled by individual households is a plausible idea. This can be achieved with rainwater harvesting (RWH) (Kahinda et al., 2007). Two kinds of RWH can be found: roof rainwater harvesting (RRWH) and ground surface rainwater harvesting (GRWH) where water is collected from roofs or ground surface runoff and stored in above-ground or underground tanks. This water can be used for potable purposes and in some cases, small scale production activities such as domestic food gardening and livestock watering (Kahinda et al., 2007, 2010; Viljoen et al., 2012). Consequently, the South African government has listed RWH as one of the strategies needed to cope with the ever-increasing water demand and to meet the Millennium Development Goals (MDGs) (Kahinda et al., 2007, 2009; Baiphethi et al., 2010; Viljoen et al., 2012).

Until recently, the importance of RWH has not been considered in water planning strategies

and policies since government relied almost exclusively on rivers and underground supplies. This is best illustrated by the proportion of people without access to municipal or borehole water, and yet have no access to harvested rainwater, while relying on river, stream, pond, dam or spring water sources (Statistics South Africa, 2010). However, the increased levels of river pollution forced a shift towards alternative sources of water (Obi et al., 2002; Lin and Biyela, 2005).

The use of harvested rainwater for domestic purposes and consumption offers an important supplement to existing traditional water sources. Rainwater harvesting presents an attractive option for households in remote villages or areas with limited access to water supply infrastructure. This water source is also a major source of drinking water during the rainy season especially in KwaZulu-Natal and the Eastern Cape (Kahinda et al., 2007; Baiphethi et al., 2009; Denison and Wotshela, 2009). The practice of RWH has been spreading in rural South Africa, especially with the financial assistance provided by the Department of Water Affairs and Forestry to farmers with limited financial resources (De Lange, 2006). The cost of operation and maintenance of a household rainwater tank is minimal. Hence, a once off capital investment by the state to install a tank can provide a poor family with a high degree of water security and independent responsibility. In this context, RWH also enables production of vegetable and food crops to address hunger and malnutrition which is rife in rural South Africa (Kahinda et al., 2007; Stockholm Environment Institute (SEI), 2009; Everson et al., 2011; Viljoen et al., 2012).

### **2.3 Rainwater harvesting systems**

The RWH technology currently practiced in South Africa ranges from rudimentary systems comprised of makeshift gutters to sophisticated ones with multiple filtration devices. In all the cases there are three main components: the catchment surface, the delivery system and a storage facility (Figure 2.1) (Kahinda et al., 2007, 2010; Denison and Wotshela, 2009). Depending on the amount of money invested in the system, components such as first flush diverters and filters may be included to reduce debris entering the tank and manage water quality (Kahinda et al., 2007; Garrison et al., 2011).

Rainwater harvesting system designs generally consider social, technical and water demand assessment. In social assessment, the community's current rainwater harvesting practices and their opinions regarding how much they can afford to spend on a RWH system are assessed (Viljoen et al., 2012). In technical assessment studies, rainfall data, existing water sources, availability of construction materials and type, as well as size of potential harvesting surface areas and taken into consideration (Kahinda et al., 2007; Garrison et al., 2011). The technical challenge in assessing the viability of RWH with respect to both domestic and gardening purposes have greatly been lessened in South Africa by the use of the computer program, SAPWAT3. This computer program is extensively applied for decision-making procedure in the estimation of crop irrigation requirements. The SAPWAT3 includes a RWH module, which is aimed at small farms or household food gardens with irrigated areas of less than one hectare. The programme uses 50 year-daily weather records provided by the derived weather stations. This is critical for the development of viable RWH systems since a thorough understanding of rainfall pattern is essential. The SAPWAT3 includes a RWH module which incorporates both irrigation and household requirements (van Heerden et al., 2009).



Figure 2. 1: A typical roof rainwater harvesting system.

### 2.3.1 Catchment surface

The catchment area of a RWH system is the surface, which receives the rainfall and channels the water to the storage tank. The most common catchment surfaces for rainwater harvesting are the roofs of dwellings, courtyards, threshing areas, paved walking areas and plastic

sheeting (van Heerden et al., 2009). In South Africa, most houses in rural areas and informal settlements have corrugated iron sheet roofing, making it the most common type of RWH surface (Kahinda et al., 2007). The roofs can either be painted red or green and some may not be painted or can be extensively corroded. Water quality from different catchment surfaces is a result of the type and state of roof material, climatic conditions, and the environment (Houston and Still, 2002).

Rainwater is generally acidic in nature and can react with the roof and compounds retained therein or by the roof, causing them to leach out (Mendez et al., 2011). Furthermore, roof temperatures are usually high due to direct solar radiation and little shading. This may accelerate chemical reactions of materials and compounds that have accumulated on roofs (Zhu et al., 2004; Mendez et al., 2011; Lee et al., 2012). Combining the constituents from roofs with elements from precipitation, chemical decomposition, and acid leaching make the quality of roof runoff questionable (Kim et al., 2005; Evans et al., 2006; Martin et al., 2007; Kahinda et al., 2007; Mendez et al., 2011; Vialle et al., 2011; Kaushik et al., 2012). When rain falls, the water dissolves particulate matter in the atmosphere and when it comes in contact with the roof the water quality may be worsened depending on the kind of materials used for roof cover, the drainage systems and their interaction with the atmospheric deposition. The variation of roof runoff quality seems to reflect differences in roofing materials, age and management, the surrounding environment, season, rainfall duration and intensity, and air quality conditions of the region (Radaideh et al., 2009; Kaushik et al., 2012).

Little information is available on the influence of painting or corroded roof surfaces on harvested rainwater quality in South Africa and the world at large. Furthermore, the proximity of animal housing to homesteads means that faecal material may be blown onto roof surfaces and washed into tanks. Only research in these areas can give the actual levels of risk and probable mitigation strategies.

### **2.3.2 Gutters**

Rainwater from the roof surface flows into gutters, pipes and then into the storage tank. A

wide variety of shapes and forms of gutters are used, including factory made polyvinyl chloride (PVC) pipes, canvas and folded metals (Sivanappan, 2006; Mendez et al., 2011; Lee et al., 2012). Half-round PVC pipes, seamless aluminium, and galvanised steel are the most common materials used for gutters (Houston and Still, 2002; van Heerden et al., 2009). Gutters are a critical part in influencing the affordability of RWH systems and the quality of the harvested water. Generally, gutters should be cheap, easy to produce, align and install, efficient in capturing runoff water, resistant to damage and easy to clean (Houston and Still, 2002). Since rainwater is slightly acidic, materials such as lead which can easily dissolve into slightly acidic water must not be used as gutter solders (Mendez et al., 2011)

Low-cost gutters can be made from waterproof canvas which is a better material than plastic as it has a longer life span and is more resistant to degradation by the sun than ordinary plastic sheeting (Houston and Still, 2002). The cleaning of gutters is critical in influencing harvested rainwater quality. Leaves, debris, dust and bird droppings may easily accumulate on the roof and gutters and can be washed off into the storage tank (Houston and Still, 2002; Amin et al., 2009; Baguma et al., 2010).

The challenge with washing of gutters is getting communities to clean gutters when they have been drinking the water without major problems for years. Given the dry spells between rain events, gutters would need to be cleaned between rain intervals, taking into account that dust blown in rural communities is mixed with animal faecal matter as animals are housed close to homesteads. Furthermore, gutter cleaning itself requires ladders, which the communities may not be able to afford and given the worn-out nature of some roofs among poor communities, cleaning gutters may present a physical danger.

### **2.3.3 First flush diverters**

During dry and windy periods, dust particles and debris are deposited on the roof and gutters. When rain falls, a high proportion of roof pollutant load will be washed into the water system and is referred to as the first flush. Exposure to ultraviolet (UV) light, heat, and desiccation on the roof destroys much of the bacteria, while wind removes some heavy metals accumulated from atmospheric fallout (Mendez et al., 2011; Kaushik et al., 2012). Pollutant

additions to roof runoff include organic matter, inert solids, faecal deposits from animals and birds, trace amounts of some metals, and even complex organic compounds (Amin and Alazba, 2011). A first flush diverter therefore ensures that runoff from the first spell of rain is flushed out and does not enter the system. This ensures to a certain degree that water quality of HRW will be better (Mendez et al., 2011). Leaf screens remove the larger debris, such as leaves, twigs and blooms that fall on the roof, while first flush diverters prevent smaller contaminants such as dust, pollen and faecal droppings from entering the RHW holding system (Virginia Department of Health, 2011).

Although a number of rural households in South Africa have installed RWH tanks, most are not equipped with first flush diverters. The challenge in rural communities is low rainfall and households are not prepared to lose any amount of rainfall they can potentially harvest. The question then is if first flush diverters are integrated in rural RWH systems, will this technology be adopted to contribute to better water quality.

#### **2.3.4 Storage tanks**

There are various options available for HRW storage tanks with respect to the shape, size and the material of construction. Storage containers include mortar jars, ferro-cement jars, and cylinders and cuboids of plastered brick, steel drums, corrugated iron cylinders, reinforced concrete tanks, and plastic drums. Storage tanks offer a range of beneficial and natural treatment processes to improve the quality of the water while it is stored (Stockholm Environment Institute, 2009). The quality of rainwater stored in a tank can potentially improve with time. Many of the organic compounds, heavy metals and other contaminants may adsorb onto suspended solids and precipitate out of the water column, and settle at the bottom of the tank forming sediments that can be removed from the water body (Amin and Alazba, 2011). Most of the organic substances in water are biodegraded to varying extents and will therefore disappear with time. Different bacteria and fungi can conglomerate in a macro-layer to form biofilms on the inside tank lining. Biofilms have been shown to adsorb heavy metals, organics, and pathogens from the water (Schets et al., 2010). It is also important to note that older biofilms may flake off into the water and re-contaminate the water source (Evans et al., 2006).

Households who cannot afford proper storage tanks currently use a number of makeshift containers including bathing dishes, metal and plastic drums. Metal drums can be extensively corroded with visible rust, whereas algal biofilm accumulation can rapidly develop in plastic containers. Most of these containers are open and subject to contamination compared to proper RWH tanks.

#### **2.4 The acceptability of rainwater harvesting**

Over the years, a number of individuals and organisations have advocated for RWH. However, the uptake and incorporation of RWH in water policies has been slow (Baiphethi et al., 2010; Kahinda et al., 2010; Kahinda and Taigbenu, 2011). Although the South African government has equipped a number of rural villages with systems to deliver piped municipal water, the supply is often unreliable and RWH systems are considered as an alternative option to meet household water needs (Houston and Still, 2002). Nevertheless, the lack of financial and labour resources to put up harvesting and storage systems has resulted in few households adopting this technology. Consequently, the department dealing with water affairs adopted a policy for farmers with limited financial resources through which certain government subsidies can be granted in terms of sections 61 and 62 of the National Water Act (Act 36 of 1998). The policy provides six possible opportunities (subsidy products) and one of these subsidy products is a grant for a rainwater tank for domestic food production and other small productive uses (Viljoen et al., 2012).

The slow rate of provision of water services among rural households in South Africa is a result of many factors, in particular the limitation of resources to provide large water distribution systems that can reach households spread over relatively large geographical areas (Viljoen et al., 2012). The terrain in some areas such as the Eastern Cape Province makes it very difficult to pump water to households on higher altitude (Baiphethi et al., 2010; Kahinda et al., 2010; Kahinda and Taigbenu, 2011). This leaves RWH as the best option for such areas during the rainy season. However, if RWH technology is to be used more widely, there is a need to integrate the technology in water resources management policies as a tool for improving water availability and quality in both rural and urban households (Houston and

Still, 2002).

Many households who would otherwise benefit from RWH lack the knowledge of the best techniques suited to their areas, and the resources and technical expertise for the construction and maintenance of the system (Baiphethi et al., 2010; Kahinda et al., 2010; Kahinda and Taigbenu, 2011). There is therefore a need for the provision of information in user-friendly formats to potential adopters. Agricultural extension and community development agents need to be equipped with the necessary skills to enable them to promote the adoption of RWH and provide technical advice to households.

## **2.5 Legal aspects of rainwater harvesting**

Strict application of the current water legislation in South Africa; the Water Services Act (Act No. 108 of 1997) and the National Water Act (Act No. 36 of 1998) as reviewed by Kahinda et al. (2009), makes RWH illegal. This is mainly because the law does not give a clear legal framework for RWH adoption. However, RWH is a non-licence, permissible water source under the National Water Act, section 22 (1) (c) (Schedule 1). On the contrary, the Water Services Act, Section 6 (1) stipulates that RWH requires authorisation from the water services authority.

The DWAF RWH programmes currently in use to improve access to clean water and food production in poor rural communities are implemented under a sub-directorate of the DWAF directorate, (Water Resources Finance and Pricing) which falls under the Agriculture Water Use Development Finance. The financial assistance to install RWH systems in communities is provided by the government in terms of sections 61 and 62 of the National Water Act (Act No. 36 of 1998). This Act empowers the Minister to avail various types of subsidies or grants to improve communities livelihoods (Kahinda et al., 2007; Everson et al., 2011).

## **2.6 The potential impact of rainwater harvesting on water related diseases**

In 2003, after experiencing the nationwide cholera epidemic of 2001-2002, the South African government released a formal statement recognizing that lack of access to safe water and



sanitation limits opportunities available for disadvantaged communities to escape poverty and worsens health risks of vulnerable groups (Hemson et al., 2006). Hence, ensuring access of the poor to sustainable, affordable and adequate levels of defined basic water supplies and sanitation services became a key focus of South Africa's water services policy (Department of Water Affairs and Forestry, 1997). Nevertheless, many households in rural areas still depend on unclean water sources such as river, dam or pool water (Statistics South Africa, 2010). The impact of water-borne diseases in South Africa is significant. It is estimated that diarrhoeal disease accounts for 21% of deaths among children aged 1–5 years (Nannan et al., 2012).

Rainwater harvesting has the potential to supply water of better quality at household level, therefore reducing the risk of water related diseases. Rainwater harvesting has not been widely utilised for drinking purposes where piped water exists due lack of information on the presence and risk from chemical and microbiological pollutants. Furthermore, there is lack of appropriate guidelines that specify the use of HRW for both potable and non-potable purposes and how the risk from chemical and microbiological pollutants can be managed (Gómez-Duarte et al., 2010; Ahmed et al., 2011; Dobrowsky et al., 2014). However, despite quality and safety issues RRWH may be the only safer option of potable water in some poor communities. Nevertheless, in the South African context, water quality takes another dimension, considering that an estimated of 5.5 million people are infected with human immune deficiency virus (HIV), one of the worst epidemic infection rates in the world (Simbayi et al., 2007).

## **2.7 Roof harvested rainwater quality**

The quality of rainwater harvested from roofs is generally considered to be safe and to meet the standards of international guidelines for drinking water quality (Zhu et al., 2004). However, other studies have reported the presence of chemical and/or microbial contaminants at levels exceeding international guidelines of drinking water (Gómez-Duarte et al., 2010; Ahmed et al., 2011; Dobrowsky et al., 2014). Differences in findings on the quality of HRW are a result of the link between specific pollution factors and the environment. The quality of the HRW depends on the characteristics of the area. These include topography, proximity to

pollution sources, weather conditions, type of catchment area and storage water tank and, management and handling practices (Ahmed et al., 2011). Therefore, legitimate questions have arisen from health regulators regarding the quality of water and consequent public health risks. Hence, the most significant issue in relation to RRWH include concerns regarding direct health implications due to chemical and microbial contaminants. Furthermore, there are direct and indirect health implications due to insect and parasite vector breeding related to water storage, for example mosquitoes and malaria disease (Kahinda et al., 2007).

Water related diseases of concern to RRWH can be classified into water-borne, based and related diseases (Bradley, 1974). Water-borne diseases are caused by the ingestion of water contaminated with faecal matter, containing pathogenic bacteria or viruses, whereas water based diseases are caused by parasites found in intermediate hosts living in water i. e. schistosomiasis and water related diseases are those transmitted by insect vectors which breed in water (Howard et al., 2006; Kahinda et al., 2007). A number of pathogens have been reported to contaminated RHRW including the bacteria; *Campylobacter* spp., *Listeria* spp., *Salmonella* spp., *Aeromonas* spp., *Vibrio* spp., *Yersinia*, spp. and *Escherichia coli* O157:H7, the protozoa *Giardia* spp. and *Cryptosporidium* spp., as well as *E. coli*, faecal coliforms and enterococci bacteria used as pollution indicators (Ahmed et al., 2011).

Rainwater can either be harvested from roof runoff or ground surface runoff. Water harvested using the latter method has contamination pathways similar to those of natural water bodies. In the case of rainwater harvested from roof surfaces, a wide array of pathogens and chemical contaminant could be present in dust, soil, nearby factories or mining activities, tree leaves, chemical deposits from nearby pesticide sprayed fields, and faeces of birds, insects, mammals and reptiles that have access to the roof. Consequently, following rain events, animal droppings, other inorganic and organic debris and chemical deposits on the roof and gutter can be transported into the tank via roof runoff. In this scenario, if the untreated water collected from the roof is used for drinking, there is a potential health risk when people use the water (Ahmed et al., 2011). However, RWH provides opportunities for government and industry to address societal needs and the many issues around poverty reduction and rural communities' constraints (Table 2.1). In order to reduce the risk of HRW contamination a

number of RWH management practices can be followed. These include cleaning of the catchment area and gutters, installation of first flush diverters and filters on tank inlet, periodical cleaning of the interior of the water storage tank and making sure it is always closed (Sazakli et al., 2007).

Table 2. 1: List of challenges and opportunities in roof rainwater harvesting systems

Challenges	Opportunities
<ul style="list-style-type: none"> <li>• Leaching of chemicals from painted surfaces and corroded roofing material.</li> <li>• Collection of debris on roof surface and gutters.</li> <li>• Absence of first flush diverters on harvesting systems.</li> <li>• Biofilm accumulation inside tanks.</li> <li>• Faecal contamination of harvested rainwater and associated pathogens.</li> <li>• Factory, mine and agricultural activities close to communities harvesting rainwater.</li> <li>• Lack of specific legislation on rainwater harvesting.</li> <li>• Ignorance among communities on the risks associated with roof harvested rainwater.</li> </ul>	<ul style="list-style-type: none"> <li>• Research into the significance of various roofing materials and their status on harvested water quality.</li> <li>• Educating communities on the maintenance of rainwater harvesting systems.</li> <li>• Integrating first flush diverters on all rainwater harvesting (RWH) systems and educating communities about their importance and correct use.</li> <li>• Research on the accumulation of biofilm in tanks, their influence on water quality and its effective removal.</li> <li>• Research on the sources of faecal contamination, implications and remediation options.</li> <li>• Research on the influence of mine dust, pesticides on roof harvested rainwater quality and evaluation of potential health risks.</li> <li>• Integrating RWH systems in government planning and policies.</li> <li>• Educating communities on contamination aspects in RWH, associated health risks and effective intervention strategies.</li> </ul>

## 2.8 Conclusion and recommendations

The South African government has made great progress in the provision of piped water to communities, especially close to towns and cities (Kahinda et al., 2007). However, provision of piped water to rural communities that are distant from towns and cities, still remains a challenge. A significant number of households in rural communities therefore still rely on river, pond or dam water for potable purposes (Kahinda et al., 2007). Although these water

sources have been used over the years, they are considered not safe as they are prone to extensive faecal contamination. People who are continuously exposed to microbial contaminated water may become immune to most of the pathogens. However, given the high levels of HIV among these communities the presence of potentially pathogenic bacteria in the water is of critical concern (UNAIDS, 2006). This is because HIV makes the people more susceptible to diseases. Furthermore, the cholera (Mugero and Hoque, 2001; Hemson et al., 2006) and typhoid (Sidley, 2005) incidences serve to show the susceptibility to water-borne diseases within these communities.

In an attempt to mitigate these risk factors the government has embarked on distribution of RWH systems to communities without access to piped water (Department of Water Affairs and Forestry, 2012). Rainwater harvesting has traditionally been practiced by a number of households in these communities without any documented reports of diseases emanating from its use (Kahinda and Taigbenu, 2011). It cannot be ignored that there is a body of evidence which undoubtedly shows the link between diseases associated with the use of HRW (Ahmed et al., 2011).

It is imperative that community-based workers should be trained on how to advise rural communities in the maintenance of RWH systems. The knowledge package should incorporate cultural norms and practices in RWH and handling within the communities. The major challenge in most of these communities is animal rearing traditions and the presence of animal faeces in close proximity to the households. This may be in the form of animal housing, of which for security reasons are built near households.

Research on harvested rainwater in rural settings such as the ones in South Africa is limited (Gómez-Duarte et al., 2010; Dobrowsky et al., 2014). Data from other situations cannot be directly extrapolated to rural settings. There is therefore a need to institute research on the actual levels of risk in these communities. The focus of these studies should not only target faecal indicators but also the presence and type of pathogens to provide a factual overview of the level of risk. Various environmental settings where rainwater harvesting is practiced should be taken into account. Furthermore, water sources that have traditionally been used by these communities should be integrated into a holistic water security program for the people.

Another important factor to consider is the availability of water based on seasonality. In most regions of South Africa, rural communities will remain dependant on alternative water sources for a sustainable water supply. In addition to the provision of RWH tanks, the communities should be equipped with information on managing the alternative water sources.

There is thus a need to gather data on the reliability of the various water sources including chemical and microbiological safety. The data should then be used to analyse trends and develop appropriate management strategies and treatment systems. Close cooperation between government, scientists, engineers and social workers is therefore needed. No single solution will be sufficient to solve rural water challenges given the diverse nature of the problem within various community environmental settings. Solutions should therefore be specific to the needs of the people given their environmental setting, affordability and sustainability.

## 2.9 References

- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Amin, M. T., and A. A. Alazba. 2011. Probable sources of rainwater contamination in a rainwater harvesting system and remedial options. *Australian Journal of Basic and Applied Sciences* 5: 1054–1064.
- Amin, M. T., M. Han, T. Muhammad, and H. Mooyoung. 2009. Probable sources of microbial contamination of stored rainwater and its remediation. In the 2nd International Conference On Rainwater Harvesting and Management, 7 - 11 September. Tokyo, Japan.
- Baguma, D., W. Loiskandl, I. Darnhofer, H. Jung, and M. Hauser. 2010. Knowledge of measures to safeguard harvested rainwater quality in rural domestic households. *Journal of Water and Health* 8: 334–345.
- Baiphethi, M. N., M. F. Viljoen, and G. Kundhlande. 2008. Rural women and rainwater harvesting and conservation: anecdotal evidence from the Free State and Eastern Cape Provinces. *Agenda: empowering woman gender equality* 22: 163–171.

- Baiphethi, M. N., M. F. Viljoen, G. Kundhlande, and N. G. Ralehlolo. 2010. Enhancing food and livelihood security in the context of the food and financial crisis: challenges and opportunities for small scale rainwater harvesting and conservation. In the Joint AAAE Third Conference/AEASA 48th Conference, 19-23 September, Cape Town, South Africa (No. 96640). African Association of Agricultural Economists (AAAE) and Agricultural Economics Association of South Africa (AEASA).
- Bradley, D. J. 1974. Water supplies: the consequences of change. In: Ciba Foundation Symposium 23 Human Rights in Health, pp 81–98.
- Cothren, P. 2013. Understanding rural water supply and access in South Africa. Msc Thesis. North Carolina State University.
- Cottle, E., and H. Deedat. 2002. The cholera outbreak: A 2000-2002 case study of the source of the outbreak in the Madlebe Tribal Authority areas, uThungulu Region, KwaZulu-Natal. Durban. Health Systems Trust.
- De Lange, M. 2006. War on hunger: some impacts of the Department of Water Affairs and Forestry rainwater harvesting pilot programme. Report to the Water Resources Finance and Pricing DWAF directorate. Pretoria, South Africa.
- Denison, J., and L. Wotshela. 2009. Indigenous water harvesting and conservation practices: Historical context, cases and implications. WRC Report No. TT 392/09. Water Research Commission (WRC). Pretoria.
- Department of Water Affairs and Forestry. 2012. Address by the Deputy Minister of Water and Environmental Affairs, Ms Rejoice Mabudafhasi. National Assembly. Budget Vote Number 38. DWAF, Pretoria, South Africa.
- Department of Water Affairs and Forestry. 1997. RDP Rural Water Supply Design Criteria Guidelines. Pretoria, South Africa.
- Dobrowsky, P. H., A. van Deventer, M. De Kwaadsteniet, T. Ndlovu, S. Khan, T. E. Cloete, and W. Khan. 2014. Prevalence of virulence genes associated with pathogenic *Escherichia coli* strains isolated from domestically harvested rainwater during low- and high-rainfall periods. Applied and Environmental Microbiology 80: 1633–8.

- Evans, C. A., P. J. Coombes, R. H. Dunstan, T. Harrison, A. Martin, and J. N. Harris. 2006. Coliforms, biofilms, microbial diversity and the quality of roof harvested rainwater. *Water Research* 40: 34–44.
- Everson, C., T. M. Everson, A. T. Modi, D. Csiwila, M. Fanadzo, V. Naiken, R. M. B. Auerbach, M. Moodley, S. M. Mtshali, and R. Dladla. 2011. Sustainable techniques and practices for water harvesting and conservation and their effective application in resource- poor agricultural production participatory adaptive research through participatory adaptive research. WRC Report No. 1465/1/11. Water Research Commission, Pretoria.
- Garrison, N., N. Resources, D. Council, J. Devine, and S. David. 2011. Capturing Rainwater from rooftops: An efficient water resource management strategy that increases supply and reduces pollution. Natural Resources Defense Council (NRDC) New York.
- Gómez-Duarte, O. G., J. Bai, and E. Newel. 2010. Detection of *Escherichia coli*, *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica*, *Vibrio cholerae*, and *Campylobacter* spp. enteropathogens by Three-reaction Multiplex PCR. *Diagnostic Microbiology and Infectious Disease* 63: 1–9.
- Hemson, D., B. Dube, T. Mbele, R. Nnadozie and D. Ngcobo. 2006. Still paying the price: revisiting the cholera epidemic of 2000-2001 in South Africa. . Municipal services project (MSP). Grocott's Publishers and Printers, Grahamstown.
- Houston, P. C., and D. A. Still. 2002. Rain water harvesting: A neglected rural water supply option. In: Biennial Conference of the Water Institute of Southern Africa (WISA). 19 – 23 May 2002, Durban, South Africa.
- Howard, G., M. F. Ahmed, A. J. Shamsuddin, S. G. Mahmud, and D. Deere. 2006. Risk assessment of arsenic mitigation options in Bangladesh. *Journal of Health Population and Nutrition* 24: 346–355.
- Kahinda, J. M., A. E. Taigbenu, and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth, Parts* 32: 1050–1057.
- Kahinda, J. M., A. E. Taigbenu, and R. J. Boroto. 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.

- Kahinda, J. M., A. E. Taigbenua, E. S. B. Lillieb, and R. J. Boroto. 2009. A GIS-based decision support system for rainwater harvesting (RHADESS). *Physics and Chemistry of the Earth Parts ABC* 34: 767–775.
- Kahinda, J. M., and A. E. Taigbenu. 2011. Rainwater harvesting in South Africa: Challenges and opportunities. *Physics and Chemistry of the Earth Parts ABC* 36: 968–976.
- Kaushik, R., R. Balasubramanian, and A. A. de la Cruz. 2012. Influence of air quality on the composition of microbial pathogens in fresh rainwater. *Applied and Environmental Microbiology* 78: 2813–2818.
- Kim, R. H., S. Lee, Y. M. Kim, J. H. Lee, S. K. Kim, and S. G. Kim. 2005. Pollutants in rainwater runoff in Korea: their impacts on rainwater utilisation. *Environmental Technology* 26: 411–420.
- Lee, J. Y., G. Bak and M. Han, and J. Y. Lee. 2012. Quality of roof harvested rainwater-comparison of different roofing materials. *Environmental Pollution* 162: 422–429.
- Lehloesa, L. J., and N. Y. O. Muyima. 2000. Evaluation of the impact of household treatment procedures on the quality of groundwater supplies in the rural community of the Victoria District, Eastern Cape. *Water SA* 26: 285–290.
- Lin, J., and P. T. Biyela. 2005. Convergent acquisition of antibiotic resistance determinants among the Enterobacteriaceae isolates of the Mhlathuze River, KwaZulu-Natal, South Africa. *Water SA* 31: 257–260.
- Martin, A. R., P. J. Coombes, R. H. Dunstan, C. A. Evans, and A. C. Morrow. 2007. The passage of direct precipitation to rainwater storage: A case study; Newcastle, Australia. In *IRCSA Conference on 21-27 Aug 2007*. Sydney Australia.
- Mendez, C. B., J. B. Klenzendorf, B. R. Afshar, M. T. Simmons, M. E. Barrett, K. A. Kinney, and M. J. Kirisits. 2011. The effect of roofing material on the quality of harvested rainwater. *Water Research* 45: 2049–2059.
- Mugero, C., and A. K. M. Hoque. 2001. Review of cholera epidemic in South Africa, with focus on KwaZulu-Natal province. Provincial Department of Health, KwaZulu-Natal.
- Mukheibir, P. 2008. Water resources management strategies for adaptation to climate-induced impacts in South Africa. *Water Resources Management* 22: 1259–1276.



- Nannan, N., R. Dorrington, R. Laubscher, N. Zinyakatira, M. Prinsloo, T. Darikwa, R. Matzopoulos, and D. Bradshaw. 2012. Under-5 mortality statistics in South Africa: Shedding some light on the trends and causes 1997-2007. South African Medical Research Council, Cape Town.
- Nevondo, T. S., and T. E. Cloete. 1999. Bacterial and chemical quality of water supply in the Dertig village settlement. *Water SA* 25: 215–220.
- Obi, C. L., and Bessong, P. O. 2011. Diarrhoeagenic bacterial pathogens in HIV-positive patients with diarrhoea in rural communities of Limpopo province, South Africa. *Journal of Health, Population and Nutrition* 20: 230-234.
- Obi, C. L., N. Potgieter, P. O. Bessong, and G. Matsaung. 2002. Assessment of the microbial quality of river water sources in rural Venda communities in South Africa. *Water SA* 28: 287–292.
- Radaideh, J., K. Al-zboon, A. Al-harashsheh, and R. Al-adamat. 2009. Quality assessment of harvested rainwater for domestic uses. *Environmental Sciences* 2: 26–31.
- Sazakli, E., A. Alexopoulos, and M. Leotsinidis. 2007. Rainwater harvesting, quality assessment and utilisation in Kefalonia Island, Greece. *Water Research* 41: 2039–2047.
- Schets, F. M., R. Italiaander, H. H. J. L. Van Den Berg, and A. M. De RodaHusman. 2010. Rainwater harvesting: quality assessment and utilisation in The Netherlands. *Journal of Water and Health* 8: 224–235.
- Sidley, P. 2005. Typhoid outbreak prompts protests over inadequate water system. *British Medical Journal* 331: 655.
- Simbayi, L. C., S. Kalichman, A. Strebel, A. Cloete, N. Henda, and A. Mqeketo. 2007. Internalised stigma, discrimination, and depression among men and women living with HIV/AIDS in Cape Town, South Africa. *Social Science and Medicine* 64: 1823–1831.
- Sivanappan, R. K. 2006. Rain water harvesting, conservation and management strategies for urban and rural sectors. In *National Seminar on Rainwater Harvesting and Water Management* 11-12 November. 2006, Nagpu.
- Statistics South Africa. 2010. General household survey. Statistical Release, May 2010 Pretoria.

- Stockholm Environment Institute (SEI). 2009. Rainwater harvesting: A lifeline for human well-being. Job No. DEP/1162/NA UNEP. United Nations Environment Programme.
- United Nations Programme on HIV/AIDS (UNAIDS). 2006. Report on the global AIDS epidemic. Geneva: UNAIDS.
- Van Heerden, P. S., C. T. Crosby, E. Theron, R. E. Schulze, and M. H. Tewolde. 2009. Integrating and updating of SAPWAT and PLANWAT to create a powerful and user friendly irrigation planning tool. WRC Report No. TT 391/08. Pretoria, South Africa.
- Vialle, C., C. Sablayrolles, M. Lovera, S. Jacob, M. -C. Huau, and M. Montrejaud-Vignoles. 2011. Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. *Water Research* 45: 3765–3775.
- Viljoen, M. F., G. Kundhlande, M. N. Baiphethi, P. Esterhuyse, J. J. Botha, J. Minkley, and G. A. Anderson. 2012. An assessment of the social and economic acceptability of rainwater harvesting and conservation practices in selected peri-urban and rural communities. WRC Report No. 1648/1/12. Pretoria, South Africa.
- Virginia Department of Health. 2011. Virginia Rainwater Harvesting and Use Guidelines. Virginia Department of Health, USA.
- Webster, M. 1999. Effective demand for rural water supply in South Africa. In proceedings of the WEDC conference (vol 25. pp 426-429). Loughborough University, Leicestershire, UK.
- World Health Organisation (WHO). 2008. Millennium development goals. New Delhi: WHO Regional Office for South-East Asia.
- Zhu, K., L. Zhang, W. Hart, M. Liu, and H. Chen. 2004. Quality issues in harvested rainwater in arid and semi-arid Loess Plateau of northern China. *Journal of Arid Environments* 57: 487–505.

## CHAPTER THREE

### STRATEGY TOWARDS THE DEVELOPMENT OF RAINWATER HARVESTING RISK ASSESSMENT AND QUALITY GUIDELINES

#### **Abstract**

Harvested rainwater (HRW) is one of the most effective alternative water sources for potable and non-potable purposes in rural communities without access to municipal water. However, HRW is prone to microbial contamination by several zoonotic bacterial and protozoan pathogens. The water is generally considered to be clean and is used without prior treatment; hence, there is a potential public health risk associated with microbial pathogens. This paper describes a systematic approach on the use of microbial risk assessment (MRA) as a tool for the management of health risks associated with HRW. Implementation of the MRA process is described as a systematic approach to determine the level of potential health risk associated with microbial pathogens in HRW for potable or non-potable purposes. Predetermined health targets are set in the MRA to achieve a degree of public health protection. Investigations into the microbial quality of HRW have revealed a wide range of contamination factors which vary among specific areas, sites and catchment systems. Due to this intricate and dynamic nature of the system, it is necessary to weigh each contaminant factor individually in the construction of MRA health risks models. This will bring about the effective management and mitigation of health risks associated with potable uses of HRW.

**Kew Words:** harvested rainwater quality, microbial contamination, health risk assessment.

### 3.1 Introduction

Harvested rainwater (HRW) is one of the most effective alternative water sources for potable in rural communities who do not have access to municipal water. However, HRW is susceptible to microbial contamination and several zoonotic bacterial and protozoan pathogens have been reported (Ahmed et al., 2008, 2010a). The potential of HRW contamination emanates from rainwater runoff draining of faecal matter including dust and debris from the catchment surface into the storage tank (Ahmed et al., 2011). The quality of harvested rainwater is therefore a function of the harvesting system structure, management and maintenance. Consequently, good quality harvested rainwater can only be attained by avoiding contamination during harvesting and storage through appropriate management practices.

The first step towards effective management of HRW as a water source is to make a thorough evaluation of the harvesting system, environmental and site specific factors, potential contaminants and levels of contamination that can develop (Schets et al., 2010; Daoud et al., 2011; Ahmed et al., 2012). The presence of potentially pathogenic microbes in HRW does not always lead to disease incidences in the communities; therefore, the fact that people are not sick does not indicate that the water source is clean, healthy or safe. Safety is assessed by comparing concentrations of water constituents against set standards for specific water uses. Tabulated guidelines are useful as a first level of assessment, but they tend to be conservative and may underestimate risk by not being able to include variation within a target user group and HRW specific factor (Ahmed et al., 2008; Daoud et al., 2011; Meyer and Casey, 2012).

We describe in this chapter a systematic approach on the use of microbial risk assessment (MRA) to illustrate its use in the management of the health risks associated with HRW systems. Microbial risk assessment has successfully been used as a tool to evaluate the probability of adverse effects on human health as a consequence of exposure to infectious agents (Fewtrell and Kay, 2007; van Lieverloo et al., 2007; Tromp et al., 2010). Hence, risk assessment has become an important instrument for the development of water and food safety management regulatory actions (Barker, 2010).

### 3.2 The need for risk assessment criteria and guidance information

There is an urgent need to formulate and provide risk assessment criteria and hence guidance information on rainwater harvesting (RWH) practices as a function of the total water consumed, RWH system, area specific environmental factors and, site or homestead specific factors. The strategy should entail the establishment of a reliable database of potentially pathogenic microbes which can contaminate HRW and estimate the risk of infection by applying quantitative microbial risk assessment (QMRA) analysis (Ahmed et al., 2010b). The data generated will provide a framework of associated risks, and the development of a system of interactive HRW quality guidance information for implementation in exposure and risk assessments.

Investigations into the microbial quality of roof harvested rainwater (RHRW) from several researches have revealed a wide range of factors that contribute to RHRW microbial contamination (Heyworth et al., 2006; Ahmed et al., 2010, 2011, 2012; Imteaz et al., 2011). These factors differed markedly among specific areas, sites and RHRW systems. Due to the complex situation, it is necessary to develop a scale or grid to weigh each water contaminant factor according to the RWH system, general environment and site or homestead specific factors, somewhat simplifying the interpretation and application of analytical data (Environment Agency, 2003; Heyworth et al., 2006; Ghisi, 2009).

Specific RWH guidelines relating to rural or urban water supplies are not readily available in South Africa. Hence, the Reconstruction and Development Programme (RDP) Rural Water Supply Design Criteria Guidelines are generally used (Department of Water Affairs and Forestry, 1997). These guidelines provide basic information on prevention of rainwater collection surfaces contamination, construction materials, first flush systems, storage units, harvesting methods and maintenance procedures. However, Guidelines for Human Settlement, Planning and Design (Council for Scientific and Industrial Research (CSIR), 2000), recommends that rainwater “*should be considered as a supplementary supply for non-potable use since it could pose a health risk*”. Given this scenario it is important to note that, there is need for effective management solutions for the safety assurance of HRW. Since no water supply is 100% safe all the time, the issue is more about the acceptable level of risk, based on cultural and socio-economic standards and the quality and availability of alternative water supplies (Lye, 2002; Ahmed et al., 2008, 2010b; Despains et al., 2009).

### 3.3 Microbial risk assessment of roof harvested rainwater

Roof harvested rainwater quality has been reported to be generally acceptable for use as a potable water source, although there is limited supporting evidence from epidemiological studies (Islam et al., 2010; Al-Salaymeh et al., 2011; Amin and Alazba, 2011). Since 1978, only six disease outbreaks worldwide have been documented to be associated with the consumption of RHRW (Heyworth et al., 2006; Lye, 2009). Nonetheless, potentially pathogenic microbes that can cause serious infections have been isolated from RHRW (Ahmed et al., 2009; Karim, 2010; Schets et al., 2010). Monitoring and managing the quality of RHRW for domestic consumption is therefore an essential consideration to maximise the use of a scarce natural resource, and minimise the risk of adverse effects on the health of people and communities.

A number of factors have a bearing on the quality of HRW. These include the kind of harvesting systems, roof material, gutters, storage tanks, wind, dust, surrounding vegetation, birds and domestic animals and air quality related to industrial, mining or agricultural activities (Howard et al., 2006a; Meera and Ahammed, 2006; Despins et al., 2009; Karim, 2010; Schets et al., 2010; Amin and Alazba, 2011; Ahmed et al., 2012). Monitoring and managing the quality of RHRW for domestic consumption should be applied routinely as part of the logistics management for public access and to comply with the right to clean water. The application of comprehensive guidelines and possible models to RWH will minimise the potential health risks of HRW use. In particular, computer models of the systems will enable extensive risk assessment through inclusion of a wider data range than can be accommodated in a limited set of guidelines (Heyworth et al., 2006; Magyar et al., 2007; Kus et al., 2010; Ahmed et al., 2012; Lee et al., 2012).

Standards for microbiological quality of various water types vary greatly throughout the world. In general there is a lack of agreement on microbiological standards regarding water quality guidelines and health related issues (World Health Organisation, 2004). Current existing and applied national and international guidelines are inadequate to universally determine with precision, the fitness-for-use of RHRW. Factors that affect the quality of

RHRW are unique and specific to individual environments, nature of harvesting systems and immediate surroundings (Meera and Ahammed, 2006; Baguma et al., 2010; Ward et al., 2010; Vialle et al., 2011). Routine monitoring of the microbiological quality of RHRW for all possible pathogens is not economically, technologically, or practically feasible (Ahmed et al., 2010b). As a result faecal indicators, such as faecal coliforms, *Escherichia coli* and enterococci, are used to determine the presence of pathogens (Dillaha and Zolan, 1985; Hörman et al., 2004; Ahmed et al., 2010b). However, the major limitation in using faecal bacteria as indicators is their poor correlation with the presence of pathogenic microorganisms in water (Hörman et al., 2004; Ahmed et al., 2008, 2011). The assessment of safety limits and the identification of potential pathogens are mostly based on assumptions. The existing water quality guidelines do not include site specific factors such as the presence of domestic animals, the types of animals, the state (Ahmed et al., 2008, 2011) and type of RWH systems, environment and water use related practices (Howard et al., 2007; Ahmed et al., 2008, 2011; Amin and Alazba, 2011).

Since it is impossible to universally predict probable pathogenic microorganisms contaminating RHRW, it is imperative that accurate area and site-specific guidance information be developed. Furthermore, the diversity in RHRW systems especially the ones in water scarce impoverished rural areas requires investigation into site and RWH system specific risk management so as to ensure the fitness-for-use of HRW. We describe the development of a microbial risk assessment procedure on RHRW water quality for domestic use taking into account the complex nature of potential contamination and contributing environmental factors. The risk assessment considered here excludes the evaluation of risks associated with chemical contaminant, although chemical risk assessment is recognised as equally important.

### **3.4 Systems approach: identifying potential hazards and health risks**

In South Africa, HRW is mainly used as a private source of drinking water mostly to individual households in water scarce and remote rural areas (Ahmed et al., 2008, 2011). Rainwater harvesting has also been adopted by a number of individuals to provide irrigation water for domestic food gardens (Department of Water Affairs and Forestry, 2007). The use of RHRW for drinking purposes poses the highest possible risks of exposure from probable

contaminating pathogens, compared to other domestic uses (Zhou et al., 2010; Ahmed et al., 2010, 2011). Assessment of health risks associated with the use of RHRW requires the consideration of whether a human health hazard exists and if the hazardous material is present in sufficient dose to cause illness (Ahmed et al., 2010b). The potential for microbial contamination of HRW arises from the collection and storage processes, of which the most common hazards are enteric pathogens (Sivanappan, 2006).

A range of microorganisms from multiple sources can potentially contaminate RHRW stored in domestic tanks. While most of the pathogens are harmless, the safety of RHRW depends on excluding or limiting the presence of enteric pathogens (Meera and Ahammed, 2006). Enteric pathogens include various species of bacteria, viruses and protozoa. Faecal material (droppings deposited by birds, lizards, mice, rats, etc.) on the roof surface, in gutters or in the tank itself is the major source of enteric pathogens (Meera and Ahammed, 2006). However, enteric pathogens do not grow or survive indefinitely in water environments, unless biofilms form (Kim and Han, 2011). Although less common, rainwater can also be collected in underground tanks. If not well sealed or protected against ground runoff, RHRW in these tanks can become contaminated with enteric pathogens (Sivanappan, 2006; Kahinda et al., 2007).

### **3.4.1 Microbial contaminants**

Initially, rainwater is free of human pathogenic microorganisms, but may become contaminated at surface runoff, during storage in tanks or distribution, or during maintenance or repair of the rainwater storage and distribution system (Sivanappan, 2006; Kahinda et al., 2007). The microbiological quality of stored rainwater may depend on storage conditions such as temperature and time, but also on the materials of which the storage containers are made, maintenance of the system and hygiene practice at the tap (Schets et al., 2010).

Pathogenic parasites such as *Cryptosporidium* and *Giardia* have higher environment persistence rates than faecal indicator bacteria, hence a quantitative link between them, cannot be expected. These parasites can be present in the absence of faecal indicators (Hörman et al., 2004). Depending on materials used, maintenance status and storage time, pathogens such as *Aeromonas* may grow in the reservoirs (Havelaar et al., 1990) and when biofilms are formed, *Legionella* may colonise the system (Berry et al., 2006; Simmons et al.,



2008). Biofilms formed in water distribution systems have been reported to protect and support bacterial re-growth and to deplete disinfection agents leading to public health problems (September et al., 2007).

### 3.4.2 Zoonosis

Zoonoses are infections that are directly or indirectly transmittable between humans and animals, for example by consuming contaminated foodstuffs, water or through contact with diseased animals (Alemayehu, 2012). The seriousness of these illnesses in humans ranges from mild symptoms to severe, life-threatening conditions. Current findings on human infectious diseases suggest that about one third to one half have a zoonotic origin (Chugh, 2008). It is estimated that 75% of the new diseases that have affected humans over the past 10 years (such as the West Nile Virus), are suspected to have originated from animals or products of animal origin (Reed et al., 2003; Alemayehu, 2012). Zoonoses can be transmitted between animals and humans in a number of ways. In the case of water-borne zoonotic pathogens, transition occurs through consumption of contaminated drinking water or food. Infectious zoonotic agents in HRW include viruses such as hepatitis A virus or norovirus, bacteria such as *Campylobacter* and *Salmonella*, and parasites such as *Giardia* and *Trichinella* (Alemayehu, 2012).

Emerging infectious diseases are those for which the incidence in humans has increased in the past two decades or threatens to increase in the near future (Alemayehu, 2012). Re-emerging bacterial zoonotic diseases are those diseases which are known but reappear after a decline in incidence (Chugh, 2008). The main factors responsible for the emergence and re-emergence of those bacterial zoonotic diseases are an increased risk of exposure to zoonotic agents, the breakdown of the host's defences, the emergence of bacterial strains resistant to antibiotics and their wide spread distribution and conjunctive causes associated with the action or inaction of man (Alemayehu, 2012). Emerging or re-emerging zoonotic diseases are usually a result of increased exposure of humans to potentially pathogens commonly through food or potable water. Roof harvested rainwater can potentially provide a link between humans and animal pathogens through faecal contamination. Therefore, there is a general need to establish factors responsible for RHRW quality and its suitability for use as an

alternative source of water. This will enable the avoidance of both human and animal deaths and the potential economic problems resulting from emerging and re-emerging bacterial zoonotic diseases as a result of contaminated RHRW.

### **3.5 Quantitative microbial risk assessment**

When developing a model to assess the risk of a specific pathogen in water, intake is the most important determinant of potential infection because of its direct link with infection rates (Howard et al., 2006b; Fewtrell and Kay, 2007; van Lieverloo et al., 2007; Ahmed et al., 2010b; Machdar et al., 2013). Water intake is determined by i) physiological age, with younger individual's water in tissues, especially muscle fibres undergoing hypertrophic development; ii) physiological state; iii) environment, for example, dry climates or extremes in the temperature-humidity index; and iv) the water content of food consumed (Food and Agriculture Organisation/World Health Organisation, 2004; Meyer and Casey, 2012). Furthermore, it is critical to consider the communities susceptibility to infection as a function of the various factors such as human-immunodeficiency virus (HIV) status, malnutrition, alcoholism and age of the infected and affected individuals. These factors generally lead to compromised immune systems with the result that infection rates and disease severity may be higher than normal (Food and Agriculture Organisation/World Health Organisation, 2004).

To establish a QMRA process for estimating the human health risk associated with defined scenarios involving exposure to specified pathogens, a four-step process is commonly employed (Gerba et al., 1996). The four steps in order are (i) hazard identification, (ii) exposure assessment, (iii) dose-response assessment, and (iv) risk characterisation. Hazard identification is achieved by collating research literature reporting the presence of specific pathogens in RHRW and accessing the presence of these pathogens in a number of water samples collected from different scenarios (Ahmed et al., 2010, 2011).

In exposure assessment, the number of pathogens in HRW and the volume ingested/inhaled by a person are estimated. To estimate the possible pathogen dose received by an individual, the likely infection routes appropriate to each pathogen must be considered. Infection may occur by ingesting (accidentally during hosing or deliberately via drinking). Another possible

route is to inhale and swallow aerosols containing these pathogens (Ahmed et al., 2010). Given these possible routes of infection, the infection risk associated should be estimated per particular scenario. These scenarios are liquid ingestion due to drinking of rainwater, accidental liquid ingestion due to garden hosing, aerosol ingestion or inhalation due to showering or hosing (Zhou et al., 2007; Kahn and Stralka, 2009). As previously stated these factors will depend on water intake which is determined by specific environmental and physiological factors (Hamilton et al., 2006; Howard et al., 2006a; Ahmed et al., 2010b). While there are estimates already in place for the different scenarios, it remains to be investigated if the same applies to the South African situation. In South Africa, some households have modern RHRW systems that have extensive filtration and consequently have little or no risk of infection. However, in rural South Africa, most RHRW systems do not include piping water into the houses, as people bath in dishes or buckets and there is little gardening, hence they have limited aerosol exposure. However, some of these areas experience high environmental temperatures resulting in people consuming a lot of water compared to those in cooler environments.

### **3.5.1 Detection of pathogens in water**

Direct routine monitoring of the microbiological quality of water for all possible pathogens is not economical, technological or practically feasible as there are hundreds of different types of pathogens that can be found in water due to faecal pollution (Ahmed et al., 2010b). Microbial quality of drinking water is therefore commonly measured by testing for faecal coliforms, *E. coli*, and enterococci as indicators of faecal contamination and surrogates for the presence of pathogenic microorganisms. Most studies therefore assess the quality of RHRW based on the numbers of faecal indicators (Dillaha and Zolan, 1985; Hörman et al., 2004). The use of faecal bacteria as indicators has the limitation of having a poor correlation with the presence of other pathogenic microorganisms in water (Hörman et al., 2004; Ahmed et al., 2010b). Indicator pathogens have traditionally been measured using culture-based methods and include heterotrophic plate count, total coliforms, faecal coliforms, *E. coli*, enterococci etc. Unfortunately, conventional culture dependant methods do not always give a true representation of the microbial composition of a community (Manz et al., 1999).

Many bacterial species have been identified to have the ability to enter a viable but non-culturable state (VBNC) where they cannot be cultured but are still ineffective. A number of microorganisms that can enter VBNC state have been identified to be human pathogens. These include: *Escherichia coli* (including EHEC strains), *Listeria monocytogenes*, *Salmonella* Typhimurium, and *Shigella flexneri* (Oliver, 2005). Cells enter a VBNC state as a response to environmental stress and can return from this survival state when they become metabolically active again. Although virulence is not expected from pathogenic cells in this state, it has been shown to return to several types of pathogens upon resuscitation (Oliver, 2005). The VBNC organisms are often encountered in natural environments and are often missed. In order to consider VBNC cells, they have to be considered *insitu* without cultivation. Various new molecular biology techniques such as quantitative PCR enable the rapid, specific, and sensitive detection of many pathogens (Pfaffl, 2004; Ahmed et al., 2008, 2010b). Hence potential pathogens in source waters that are otherwise difficult and/or laborious to quantify using traditional culture methods can now be quantified *insitu*. In general, binary PCR (presence/absence) is first used to detect the pathogens before quantitative PCR (qPCR) to determine the probable concentrations of the pathogens present (Ahmed et al., 2008).

The United States Government instituted a study to evaluate qPCR for use as a national water quality metric (Shanks et al., 2012). Inter-laboratory variability based on the measurement of enterococci and *Bacteroidales* concentrations from standardised, spiked, and environmental sources of DNA was determined using the Entero1a and GenBac3 qPCR methods, respectively. Special attention was placed on the influence of the DNA isolation step and the effect of simplex and multiplex amplification approaches on inter-laboratory variability. Results indicated that inter-laboratory variability differences between protocols tested were relatively low. Major differences between laboratories were attributed to experimental error due to deviations in the execution of standardised protocols. These findings showed the importance of adopting standardised qPCR protocols, quality control metrics, and proficiency of laboratory systems and personnel for the repeatability of qPCR.

### **3.5.2 Biofilm formation in rainwater storage tanks**

Biofilms develop on biological and non-biological surfaces ubiquitous in natural aquatic environments and engineered systems. Biofilms are self-regulating communities of

microorganisms co-adhered to each other. They attach to a surface via extra-cellular polymeric substances excreted by the cells (Garrett et al., 2008). The microbial communities consist of not only bacteria but also filamentous fungi, yeasts, algae, protozoa and viruses (Lindsay and Von Holy, 2006). Biofilms can develop in the interior lining of rainwater storage tanks and on surfaces of suspended particles (Lehtola et al., 2004). A number of bacteria including *E. coli* and coliforms can be incorporated in biofilms during its development. This is particularly of concern as biofilms afford microorganisms a degree of protection from different disinfection systems that may be used (U. S. Environmental Protection Agency, 2002).

Harvested rainwater is stored in tanks for prolonged periods of time before being used. The potential of pathogen amplification within storage tanks and uptake within biofilms allowing for survival and later shedding within the system is of concern (U. S. Environmental Protection Agency, 2002). The development of biofilms in storage tanks provide reservoirs for coliforms and pathogenic microorganisms which through growth and detachment, can be responsible for the majority of contaminating pathogens in the stored water (van der Wende et al., 1989). Biofilm development is therefore of importance in RWH systems and require an effective cleaning strategy to reduce potential health risks.

### **3.6 Discussion**

The review of the published literature reported here indicates that RWH appears to be one of the most promising alternatives for supplying freshwater in the face of increasing water scarcity and escalating demand (Alcock and Verster, 1987; Zhu and Li, 2004; Schets et al., 2010; Ward et al., 2011). However, RWH sustainability requires close cooperation between government, private sector (NGOs and scientists) and the rural households. This will include an integrated system approach where the quantity/quality of the water supplied as well as the associated costs of implementation are considered (Kahinda et al., 2007; Kahinda and Taigbenu, 2011; Viljoen et al., 2012). A sustainable DRWH system is one that is implemented after considering not only the physical attributes (rainfall, location and catchment characteristics) and the socio-economic attributes in its design but also the quality of the rainwater and the alternative water sources (Kahinda et al., 2007; Stockholm

Environment Institute (SEI), 2009; Viljoen et al., 2012).

The microbial quality of RHRW has been shown to be highly variable than is commonly perceived and should be considered potentially poor unless a more rigorous microbial assessment has been undertaken (Evans et al., 2007; Schets et al., 2010; Amin et al., 2011; Daoud et al., 2011). On the basis of the reported data, the quality of RHRW appears to be strongly influenced by the season, the number of preceding dry days, animal activities in close proximity to the roof and rainwater tanks, geographical location, and other factors (Evans et al., 2006; Mendez et al., 2011; Kaushik et al., 2012). In addition, little information is currently available on the number of microbial pathogens that can be present in RRWH (Ahmed et al., 2008, 2010, 2011, 2012). Studies in literature that have assessed the quality of RHRW were mainly based on the presence or absence of particular pathogens, with scanty data on the quantitative values of the pathogens in HRW (Evans et al., 2007; Afshar et al., 2009; Amin et al., 2009, 2011). In addition, no information is available concerning the ongoing prevalence of different pathogens in HRW over time (Ahmed et al., 2011). It is recommended, therefore, that further research be undertaken on the occurrence and numbers of potential pathogens in RRWH in a range of locations over time using appropriately designed longitudinal sampling schemes. In addition, any microbial assessment should involve the analysis of RRWH for actual pathogenic species, not just the common faecal indicator bacteria.

Available data in the literature shows that the commonly used indicators such as faecal coliforms and *E. coli* may not be suitable to indicate the risk of illness from untreated RHRW due to their observed poor correlation with pathogens (Ahmed et al., 2008, 2010b). In addition, several zoonotic bacterial and protozoan pathogens capable of causing infections in humans have been detected in individual RWH systems, probably from wild or domestic animals that have access to roofs (Afshar et al., 2009; Ahmed et al., 2010, 2012). Hence, more testing of RHRW is needed to quantify the potential impact of pathogens originating from animals and the potential health risks from the different uses of this water.

A number of control studies have established links between gastroenteritis and consumption of untreated RHRW. However, these reported outbreaks tended to involve small numbers of

individuals and the reported illnesses were often related to communal RWH systems (Ahmed et al., 2011). Other studies could not identify RHRW as a source of infection and therefore could only hypothesize that RHRW was the possible source of infection via circumstantial evidence (Koplan et al., 1978; Merritt et al., 1999). It should be noted that most of these case control studies used culture-based methods to establish a link between RHRW and faecal specimens from patients. Additional studies that focus on the collection and matching of pathogenic strains from faecal specimens from self-reported incidences of gastroenteritis and from potential sources such as tap water and tank water using sensitive molecular typing methods, would provide valuable information to determine if there is a direct link between gastroenteritis and RHRW. This could be more practical than a comparable epidemiological study due to the complexity and costs of epidemiological studies and the fact that most incidences of gastroenteritis remain unreported. The use of a QMRA health risk analysis using data on the reported incidences of microbial pathogens in RHRW and a series of potential exposure pathways would be another valuable tool to assess overall health risks associated with RHRW and the need for treatment before certain uses (i. e. drinking).

It is evident that new or improved guidelines and policies for RWH systems are needed to assist in maximising the uptake and use of RWH for a range of end users while ensuring that any associated health risks are minimised through appropriate management procedures. Proper design and maintenance must be achievable to ensure the long-term success of RWH integration in rural communities and urban water systems. This should include on-going monitoring of RWH systems and health risks assessments such as those using QMRA model, and appropriate management requirements to mitigate these risks. This will only be achieved through the collection of more data on the incidence and on-going prevalence of microbial pathogens in RWH systems.

To increase the water security of rural households, all potential water sources should be evaluated with respect to number, location, yield, dependability, and quality. Because of the unreliable rainfall, it is necessary to investigate the existence of alternative water sources near a particular site that can be developed to ensure water supplies that are more reliable. For years, untreated roof runoff has been widely used for drinking purposes with very few recorded or reported serious health problems. Rainwater collected from ground catchment

systems is generally subject to higher levels of microbial contamination and has to be treated before being consumed. It is necessary when designing the RWH systems to prevent or minimise breeding of mosquitoes and avoid or reduce pollution and contamination. Guidelines for the operation and maintenance of RWH systems need to be written and disseminated to rural communities. Water legislations should be clear on RWH to facilitate its implementation.

An important issue is to assess the actual risk of using potentially unsafe HRW. Conventional MRAs are based on water quality analysis, which have however many drawbacks (Ahmed et al., 2010, 2011; Schets et al., 2010; Daoud et al., 2011). These include poor sampling design, the use of indicator bacteria of which has been shown to have poor correlation with pathogens, the lack of related locally tested dose-response functions, and the fact that water safety eventually depend on management practices (Schets et al., 2010; Islam et al., 2011; Masters et al., 2011; Domènech et al., 2012).

Although water quality regulations may vary between countries, two overarching approaches to MRA that influence the setup of related control systems are implemented worldwide. On one hand are the World Health Organisation QMRA based standards (WHO, 2006) and on the other are the USEPA, standards for water quality (Environmental Protection Agency (EPA), 2012). Both guidelines still refer to *E. coli* as most convenient pathogen indicator. However, while WHO (2006) promotes a step-wise approach for risk reduction based on the QMRA estimates, USEPA (2012) is based on the demanding water quality standards issued across the USA, which are mostly unachievable in many developing countries (Howard et al., 2006b).

The need for data allowing a robust and also geographically specific health risk assessment and monitoring is apparent given the gap between EPA water quality standards which are straight forward and the more complex QMRA assessment framework of WHO (WHO, 2006; US Environmental Protection Agency, 2012). Such an assessment should ideally be of an integrated nature looking beyond water quality and the irrigated produce to evaluate its consumption risk within the context of other risk factors the consumer is exposed to. QMRA



does allow such multi-risk assessments, which are of particular importance in low income settings where many risk factors occur simultaneously. Identifying the most important ones will help to prioritise interventions for improvement where capacities do not allow addressing all problem areas.

On the way, towards a robust risk assessment program, data generated from research should be used to design comprehensive risk assessment procedures across HRW and irrigated foodborne pathways for diseases potentially caused by drinking untreated water or eating of fresh/minimally processed produce. The use of indicator bacteria should be tested for relationships/correlation with the occurrence of pathogen, and if necessary, the type of indicator bacteria used can be revised. In order for QMRA methods to be relevant to specific environmental settings data collected locally on exposure, dose-response, etc., including local immunity levels should rather be used than data from developed countries. However, this can only be achieved through capacity building of local laboratories for diagnostic monitoring of pathogens so as to implement QMRA assessments epidemiologically. Databases should therefore be established to address the validity of *E. coli* as indicator for bacterial, viral and protozoan pathogens especially for QMRA methods that use *E. coli* as index organism as with the WHO guidelines. For vulnerable sub-populations, e. g. individuals with HIV, the risk and consequences of multiple vs. single infections should be ascertained, together with responses of different socials and ethnic groups. After all the data compilation, deterministic and stochastic QMRA methods should be compared and contrasted in the development of integrated risk assessments by identifying commonalities and differences between environmental impact assessment and health impact assessment approaches. This will enable the evaluation of probable interventions strategies cost effectiveness in order to enhance priority setting in program implementation by decision-makers.

### **3.7 Conclusion and Recommendations**

Given the lack of standards, knowledge of contamination and complexity of the problem with RWH there is a need to propose a set of guidance information on best practices and risks associated with harvesting rainwater. The guidance information should give a description of the chemical and biological constituents that can potentially affect water quality, their occurrence in the HRW interdependence with other constituents, and their properties. These

should also include standardised methods for measurement. For each contaminant, the guidance information on risks that are site specific should be provided in addition to generic guidance information for RHRW practices. While the guidance information may not be completely accurate, it should provide improvement over the current non-specific systems. The information should allow users to obtain information on risks more easily by explaining RHRW sampling procedures and how to interpret the results according to RHRW system, general environment, and site or homestead specific factors. This guidance information more specifically should specify the targeted water quality range in which infection or adverse effects are unlikely to occur. Current data gathered mainly from Australia suggest that even when concentrations are over the upper safety limit and likely to result in infection, diseases may not always develop in the community (Ahmed et al., 2011).

The response to risks depend on site specific factors, synergistic and antagonistic interactions between food and water contaminants, the age of the person, and the actual water intake that determines the ingestion rate of the contaminant. However, a precise way to evaluate the risk has limitations in that tools to measure RHRW specific factors may not be easily available. In this case, impact factors may have to be estimated, making it still precarious. Given this scenario, an interactive water quality guideline in the form of a computer software system that would have programmed references, by which risk assessment could be done for a specific RHRW system, environment, site or homestead specific factors and extent of exposure, should be developed (Meyer and Casey, 2012).

### **3.8 References**

- Afshar, B., M. J. Kirisits, and K. Kinney. 2009. Microbial source tracking in drinking water from rainwater harvesting. Texas Water Research Institute, United States Geological Survey, Texas.
- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Ahmed, W., A. Vieritz, A. Goonetilleke, and T. Gardner. 2010. Health risk from potable and non-potable uses of roof harvested rainwater using quantitative microbial risk assessment in Australia. *Applied and Environmental Microbiology* 76: 7382–7391.

- Ahmed, W., A. Vieritz, T. Gardner, and A. Goonetilleke. 2009. Microbial risk from rainwater tanks in South-East Queensland. *Water* 36: 80-85.
- Ahmed, W., F. Huygens, A. Goonetilleke, and T. Gardner. 2008. Real-Time PCR detection of pathogenic microorganisms in roof harvested rainwater in Southeast Queensland, Australia. *Applied and Environmental Microbiology* 74: 5490–5496.
- Ahmed, W., J. P. S. Sidhu, S. Toze, and L. Hodgers. 2012. Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland. *Applied and Environmental Microbiology* 5: 219–226.
- Alcock, P. G., and E. Verster. 1987. Investigation into unconventional sources of water for a peri- urban/rural district of KwaZulu. *South African Journal of Science* 83: 348–352.
- Alemayehu, A. 2012. Review on emerging and re-emerging bacterial zoonotic diseases. *American-Eurasian Journal of Scientific Research* 7: 176–186.
- Al-Salaymeh, A., I. A. Al-Khatib, and H. A. Arafat. 2011. Towards sustainable water quality: management of rainwater harvesting cisterns in Southern Palestine. *Water Resources Management* 25: 1721-1736.
- Amin, M. T., and A. A. Alazba. 2011. Probable sources of rainwater contamination in a rainwater harvesting system and remedial options. *Australian Journal of Basic and Applied Sciences* 5: 1054–1064.
- Amin, M. T., M. Han, T. Muhammad, and H. Mooyoung. 2009. Probable sources of microbial contamination of stored rainwater and its remediation. In the 2nd international conference on rainwater harvesting and management, 7 - 11 September. Tokyo, Japan.
- Amin, T. and M. Y. Han. 2011. Microbial quality variation within a rainwater storage tank and the effects of first flush in Rainwater Harvesting (RWH) System. *Australian Journal of Basic and Applied Sciences* 5: 1804–1813.
- Baguma, D., W. Loiskandl, I. Darnhofer, H. Jung, and M. Hauser. 2010. Knowledge of measures to safeguard harvested rainwater quality in rural domestic households. *Journal of Water and Health* 8: 334–345.

- Barker, G. C., C. Bayley, A. Cassidy, S. French, A. Hart, P. K. Malakar, J. Maule, M. Petkov, and R. Shepherd. 2010. Can a participatory approach contribute to food chain risk analysis. *Risk Analysis* 30: 766–781.
- Berry, D., C. Xi, and L. Raskin. 2006. Microbial ecology of drinking water distribution systems. *Current Opinion in Biotechnology* 17: 297–302.
- Chugh, T. D. 2008. Emerging and re-emerging bacterial diseases in India. *Journal of Biosciences* 33: 549–555.
- Council for Scientific and Industrial Research. 2000. Guidelines for Human Settlement, Planning and Design. CSIR Building and Construction Technology, Pretoria. CTP Book Printers (Pty) Ltd, Parow, Cape Town.
- Daoud, A. K., K. M. Swaileh, R. M. Hussein, and M. Matani. 2011. Quality assessment of roof harvested rainwater in the West Bank, Palestinian Authority. *Journal of Water and Health* 9: 525–33.
- Department of Water Affairs and Forestry. 2007. Programme guidelines for intensive family food production and rainwater harvesting. Department of Water Affairs and Forestry, Pretoria. South Africa.
- Department of Water Affairs and Forestry. 1997. RDP rural water supply design criteria guidelines. Pretoria, South Africa.
- Despins, C., K. Farahbakhsh, and C. Leidl. 2009. Assessment of rainwater quality from rainwater harvesting systems in Ontario, Canada. *Journal of Water Supply Research and Technology* 58: 117–134
- Dillaha, T. A., and W. J. Zolan. 1985. Rainwater catchment water quality in Micronesia. *Water Research* 19: 741–746.
- Domènech, L., H. Heijnen, and D. Saurí. 2012. Rainwater harvesting for human consumption and livelihood improvement in rural Nepal: benefits and risks. *Water and Environment Journal* 26: 465–472.

- Evans, C. A., P. J. Coombes, R. H. Dunstan, and T. Harrison. 2007. Identifying the major influences on the microbial composition of roof harvested rainwater and the implications for water quality. *Water Science Technology* 55: 245–253.
- Evans, C. A., P. J. Coombes, R. H. Dunstan, T. Harrison, A. Martin, and J. N. Harris. 2006. Coliforms, biofilms, microbial diversity and the quality of roof harvested rainwater. American Rainwater Catchment "American Rainwater Catchment Systems Association (ARCSA) Conference Presentations.
- Fewtrell, L., and D. Kay. 2007. Quantitative microbial risk assessment with respect to *Campylobacter* spp. in toilets flushed with harvested rainwater. *Water and Environment Journal* 21: 275–280.
- Food and Agriculture Organisation/World Health Organisation (FAO/WHO). 2004. Fruit and vegetables for health. Report of a Joint FAO/WHO Workshop, 1–3 September 2004, Kobe, Japan pp 1-46.
- Garrett, T. R., M. Bhakoo, and Z. Zhang. 2008. Bacterial adhesion and biofilms on surfaces. *Progress in Natural Science* 18: 1049–1056.
- Gerba, C. P., J. B. Rose, C. N. Haas, and K. D. Crabtree. 1996. Water-borne rotavirus: A risk assessment. *Water Research* 30: 2929–2940.
- Ghisi, E. 2009. Parameters influencing the sizing of rainwater tanks for use in houses. *Water Resources Management* 24: 2381–2403.
- Hamilton, A. J., F. Stagnitti, R. Premier, A. Boland, and G. Hale. 2006. Quantitative microbial risk assessment models for consumption of raw vegetables irrigated with reclaimed water. *Applied and Environmental Microbiology* 72: 3284–3290.
- Havelaar, A. H., J. F. M. Versteegh, and M. During. 1990. The presence of *Aeromonas* in drinking water supplies in The Netherlands. *International Journal of Hygiene and Environmental Medicine* 190: 236–256.
- Heyworth, J. S., G. Glonek, E. J. Maynard, P. A. Baghurst, and J. Finlay-Jones. 2006. Consumption of untreated tank rainwater and gastroenteritis among young children in South Australia. *International Journal of Epidemiology* 35: 1051–1058.

- Hörman, A., R. Rimhannen-Finne, L. Maunula., C. H. von Bonsdorff, N. Torvela, A. Heikinheimo, and M. L. Hänninen. 2004. *Campylobacter* spp., *Giardia* spp., *Cryptosporidium* spp., noroviruses, and indicator organisms in surface water in Southwestern Finland, 2000-2001. *Applied and Environmental Microbiology* 70: 87–95.
- Howard, G., M. F. Ahmed, A. J. Shamsuddin, S. G. Mahmud, and D. Deere. 2006a. Risk Assessment of Arsenic Mitigation Options in Bangladesh. *Journal of Health Population and Nutrition* 24: 346–355.
- Howard, G., S. Pedley, and S. Tibatemwa. 2006b. Quantitative microbial risk assessment to estimate health risks attributable to water supply: Can the technique be applied in developing countries with limited data? *Journal of Water and Health* 4: 49–65.
- Imteaz, M. A., A. Shanableh, A. Rahman, and A. Ahsan. 2011. Optimisation of rainwater tank design from large roofs: A case study in Melbourne, Australia. *Resources Conservation and Recycling* 55: 1022–1029.
- Islam, M. A., H. Sakakibara, M. R. Karim, M. Sekine, and Z. H. Mahmud. 2011. Bacteriological assessment of drinking water supply options in coastal areas of Bangladesh. *Journal of Water and Health* 9: 415–428.
- Islam, M. M., F. N. -F. Chou, and M. R. Kabir. 2010. Acceptability of the rainwater harvesting system to the slum dwellers of Dhaka City. *Water Science and Technology* 61: 1515–1523.
- Kahinda, J. M., A. E. Taigbenu, and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 32: 1050–1057.
- Kahinda, J. M., and A. E. Taigbenu. 2011. Rainwater harvesting in South Africa: Challenges and opportunities. *Physics and Chemistry of the Earth Parts ABC* 36: 968–976.
- Kahn, H. D., and K. Stralka. 2009. Estimated daily average per capita water ingestion by child and adult age categories based on USDA's 1994-1996 and 1998 continuing survey of food intakes by individuals. *Journal of Exposure Science and Environmental Epidemiology* 19: 396–404.

- Karim, M. R. 2010. Microbial contamination and associated health burden of rainwater harvesting in Bangladesh. *Water Science and Technology* 61: 2129–2135.
- Kaushik, R., R. Balasubramanian, and A. A. de la Cruz. 2012. Influence of air quality on the composition of microbial pathogens in fresh rainwater. *Applied and Environmental Microbiology* 78: 2813–2818.
- Kim, M., and M. Han. 2011. Composition and distribution of bacteria in an operating rainwater harvesting tank. *Water Science and Technology* 63: 1524–1530.
- Koplan, J. P., R. D. Deen, W. H. Swanston, and B. Tota. 1978. Contaminated roof-collected rainwater as a possible cause of an outbreak of salmonellosis. *Journal of Hygiene* 81: 303–309.
- Kus, B., J. Kandasamy, S. Vigneswaran, and H. K. Shon. 2010. Water quality characterisation of rainwater in tanks at different times and locations. *Water Science and Technology* 61: 429–439.
- Lee, J. Y, and G. Bak, M. Han. 2012. Quality of roof harvested rainwater--comparison of different roofing materials. *Environmental Pollution* 162: 422–429.
- Lehtola, M. J., I. T. Miettinen, M. M. Keinänen, T. K. Kekki, O. Laine, A. Hirvonen, T. Vartiainen, and P. J. Martikainen. 2004. Microbiology, chemistry and biofilm development in a pilot drinking water distribution system with copper and plastic pipes. *Water Research* 38: 3769–3779.
- Lindsay, D., and A. Von Holy. 2006. Bacterial biofilms within the clinical setting: what healthcare professionals should know? *Journal of Hospital Infection* 64: 313–325.
- Lye, D. J. 2002. Health risks associated with consumption of untreated water from household roof catchment systems. *Journal of the American Water Resources Association* 38: 1301–1306.
- Lye, D. J. 2009. Rooftop runoff as a source of contamination: a review. *Science of the Total Environment* 407: 5429–5434.

- Machdar, E., N. P. Van Der Steen, L. Raschid-Sally, and P. N. L. Lens. 2013. Application of quantitative microbial risk assessment to analyse the public health risk from poor drinking water quality in a low income area in Accra, Ghana. *Science of the Total Environment* 449: 134–142.
- Magyar, M. I., V. G. Mitchell, A. R. Ladson, and C. Diaper. 2007. An investigation of rainwater tanks quality and sediment dynamics. *Water Science and Technology* 56: 21–28.
- Manz, W., K. Wendt-Potthoff, T. R. Neu, U. Szewzyk, and J. R. Lawrence. 1999. Phylogenetic composition, spatial structure, and dynamics of lotic bacterial biofilms investigated by fluorescent in situ hybridisation and confocal laser scanning microscopy. *Microbial Ecology* 37: 225–237.
- Masters, N., A. Wiegand, W. Ahmed, and M. Katouli. 2011. *Escherichia coli* virulence genes profile of surface waters as an indicator of water quality. *Water Research* 45: 6321–6333.
- Meera, V., and M. M. Ahammed. 2006. Water quality of rooftop rainwater harvesting systems: A review. *Journal of Water Supply Research and Technology AQUA* 55: 257–268.
- Mendez, C. B., S. Bae, B. Chambers, S. Fakhreddine, T. Gloyna, S. Keithley, L. Untung, M. E. Barrett, K. Kinney, and M. J. Kirisits. 2011. Effect of roof material on water quality for rainwater harvesting systems – additional physical, chemical, and microbiological data Report by Texas Water Development Board. Austin, Texas.
- Merritt, A., R. Miles, and J. Bates. 1999. An outbreak of *Campylobacter enteritis* on an island resort, North Queensland. *Communicable Diseases Intelligence* 23: 215–219.
- Meyer, J. A., and N. H. Casey. 2012. Establishing risk assessment on water quality for livestock. *Animal Frontiers* 2: 44–49.
- Oliver, J. D. 2005. The viable but nonculturable state in bacteria. *Journal of Microbiology* 43: 93–100.



- Pfaffl, M. W. 2004. Quantification strategies in real-time PCR. In A-Z of quantitative PCR. S. A. Bustin (ed.) (pp. 87–112). International University Line (IUL). La Jolla, CA, USA.
- Reed, K. D., J. K. Meece, J. S. Henkel, and S. K. Shukla. 2003. Birds, migration and emerging zoonoses: West Nile virus, Lyme disease, influenza A and enteropathogens. *Clinical Medicine and Research* 1: 5–12.
- Schets, F. M., R. Italiaander, H. H. J. L. Van Den Berg, and A. M. De RodaHusman. 2010. Rainwater harvesting: quality assessment and utilisation in The Netherlands. *Journal of Water and Health* 8: 224–235.
- September, S. M., F. A. Els, S. N. Venter, and V. S. Brozel. 2007. Prevalence of bacterial pathogens in biofilms of drinking water distribution systems. *Journal of Water and Health* 5: 219–227.
- Shanks, O. C., M. Sivaganesan, L. Peed, C. A. Kelty, A. D. Blackwood, R. Greene, R. T. Noble, R. N. Bushon, E. A. Stelzer, J. Kinzelman, C. Sinigalliano, D. Wanless, Y. Cao, S. Weisberg, J. Valarie, C. Staley, K. H. Oshima, M. Varma, R. A. Haugland, M. R. Greene, T. Anan'eva, J. Griffith, and V. J. Harwood. 2012. Inter-laboratory comparison of real-time PCR protocols for quantification of general fecal indicator bacteria. *Environmental Science and Technology* 46: 945–953.
- Simmons, G., S. Jury, C. Thornley, D. Harte, J. Mohiuddin, and M. Taylor. 2008. A Legionnaires' disease outbreak: A water blaster and roof-collected rainwater systems. *Water Research* 42: 1449–1458.
- Sivanappan, R. K. 2006. Rain water harvesting, conservation and management strategies for urban and rural sectors. In: National Seminar on Rainwater Harvesting and Water Management 11-12 Nov. 2006, Nagpu.
- Stockholm Environment Institute (SEI). 2009. Rainwater harvesting: A lifeline for human well-being. Job No. DEP/1162/NA UNEP. United Nations Environment Programme.
- Tromp, S. O., H. Rijgersberg, and E. Franz. 2010. Quantitative microbial risk assessment for *Escherichia coli* O157:H7, *Salmonella enterica*, and *Listeria monocytogenes* in leafy

- green vegetables consumed at salad bars, based on modelling supply chain logistics. *Journal of Food Protection* 73: 1830–1840.
- U. S. Environmental Protection Agency. 2002. Health risks from microbial growth and biofilms in drinking water distribution systems. U. S. Environmental Protection Agency Office of Ground Water and Drinking Water Standards and Risk Management, Pennsylvania, Washington DC.
- U. S. Environmental Protection Agency. 2012. Microbial risk assessment guideline: pathogenic microorganisms with focus on food and water. EPA/100/J-12/001.
- Van der Wende, E., W. G. Characklis, and D. B. Smith. 1989. Biofilms and bacterial drinking water quality. *Water Research* 23: 1313–1322.
- Van Lieverloo, J. H. M., E. J. M. Blokker, and G. Medema. 2007. Quantitative microbial risk assessment of distributed drinking water using faecal indicator incidence and concentrations. *Journal of Water and Health* 5: 131–149.
- Vialle, C., C. Sablayrolles, M. Lovera, S. Jacob, M. -C. Huau, and M. Montrejaud-Vignoles. 2011. Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. *Water Research* 45: 3765–3775.
- Viljoen, M. F., G. Kundhlande, M. N. Baiphethi, P. Esterhuysen, J. J. Botha, J. Minkley, and G. A. Anderson. 2012. An assessment of the social and economic acceptability of rainwater harvesting and conservation practices in selected peri-urban and rural communities. WRC Report No. 1648/1/12. Pretoria, South Africa.
- Ward, S., F. A. Memon, and D. Butler. 2010. Harvested rainwater quality: the importance of appropriate design. *Water Science and Technology* 61: 1707–1714.
- Ward, S., S. Barr, F. A. Memon, and D. Butler. 2011. At the brink: the transition of rainwater harvesting systems in the UK. *12th International Conference on Urban Drainage* 3: 976–987.
- World Health Organisation. 2004. Guidelines for drinking water quality: (vol 1) recommendations. World Health Organisation, Geneva

World Health Organisation. 2006. Guidelines for the safe use of wastewater, excreta and greywater: Policy and regulatory aspects. World Health Organisation.

Zhou, Y., J. M. Benson, C. Irvin, H. Irshad, and Y. -S. Cheng. 2007. Particle size distribution and inhalation dose of shower water under selected operating conditions. *Inhalation Toxicology* 19: 333–342.

Zhou, Y., W. Shao, and T. Zhang. 2010. Analysis of a rainwater harvesting system for domestic water supply in Zhoushan, China. *Journal of Zhejiang University Science* 1: 342–348.

Zhu, Q., and Y. Li. 2004. Rainwater harvesting--an alternative for securing food production under climate variability. *Water Science and Technology* 49: 157–163.

## CHAPTER FOUR

### A SCOPING STUDY ON PREVALENCE OF *ESCHERICHIA COLI* AND *ENTEROCOCCUS* SPECIES IN HARVESTED RAINWATER STORED IN TANKS

#### Abstract

Rainwater harvesting (RWH) is a relatively inexpensive technology that has the potential to provide safe water in communities where conventional technologies are difficult to implement. In this scoping study, the microbiological quality of rainwater harvested from rooftops and ground surface runoff was evaluated based on the concentrations of *Escherichia coli*, faecal coliforms, enterococci and *Pseudomonas aeruginosa*. Samples were collected from 15 rooftop rainwater harvesting (RRWH) tanks, four ground surface rainwater harvesting (GRWH) tanks, three rivers and one spring water source in the Eastern Cape Province of South Africa and 14 RRWH tanks in Gauteng Province. In the Eastern Cape Province *E. coli*, enterococci and *P. aeruginosa* were detected in seven, four and eight of the 15 RRWH tanks, respectively. Enterococci were detected only from one river whereas *E. coli* and *P. aeruginosa* were detected in all three rivers and in spring water, only *P. aeruginosa* was detected. Samples from GRWH tanks were positive for *E. coli*, enterococci and *P. aeruginosa* in two, three and all of the four tanks, respectively. In Gauteng Province, *E. coli*, coliforms and enterococci were detected from six, six and nine of the 14 rainwater tanks, respectively. On average *E. coli* and enterococci were detected in 44.8% of the roof harvested rainwater (RHRW) tanks although enterococci concentrations were several times higher than those for *E. coli*. We further evaluated the significance of urban pigeons as the likely sources of contamination by isolating one hundred and fifty-six enterococci from thirty pigeon faecal samples and two hundred and eight enterococci from RHRW samples collected from Gauteng Province. Matrix-assisted laser desorption ionisation (MALDI-TOF-MS) identification of the various enterococci revealed the four species *E. faecalis* (20.5%), *E. mundtii* (20.51%), *E. faecium* (23.1%) and *E. casseliflavus* (17.3%) to be dominant in faecal samples whereas *E. casseliflavus* (34.6%) and *E. mundtii* (33.2%) were dominant in RHRW.

Key Words: Rainwater harvesting, contamination, indicator bacteria, health risks.

## 4.1 Introduction

Roof harvested rainwater (RHRW) is one of the major alternative water sources used in South African rural communities without access to piped water (Kahinda et al., 2007; Kahinda and Taigbenu, 2011). The water is generally considered to be clean and is used without prior treatment. This perception is supported by a number of studies which showed that RHRW poses no increased risk when compared to municipal piped water (Dillaha and Zolan, 1985; Heyworth et al., 2006). In contrast, a number of other studies on the microbial quality of RHRW have reported the presence of specific zoonotic pathogens in individual or communal RWH systems (Lye, 2002; Ahmed et al., 2010b; Islam et al., 2011; Mpogui and Mpogui, 2012).

The major source of pollution in RHRW is faeces of animals. Animals can carry a wide range of human gastrointestinal pathogens either being themselves diseased or as healthy carriers (Cox et al., 2005). The pathogens can be excreted in their faeces and may include the bacteria *Campylobacter* spp., *Listeria* spp., *Salmonella* spp., *Aeromonas* spp., *Vibrio* spp., *Yersinia* spp. and *Escherichia coli* O157:H7, the protozoa *Giardia* spp. and *Cryptosporidium* spp., as well as the bacteria used as pollution indicators including, *E. coli*, faecal coliforms and enterococci (Jagals et al., 1995; Sinton et al., 1998; Curtis et al., 2000; Cox et al., 2005).

Households in rural communities practice free range domestic animal rearing and faecal matter is a common feature on the ground around their homesteads. The presence of faecal matter is a significant contamination risk factor since dried faecal matter can be blown by wind onto roof surfaces. Following rain events, faecal droppings and other organic debris deposited on the roof and gutters can be transported into the tank with roof runoff. The actual level of risk from potential pathogens in RHRW can be influenced by several factors including the type and numbers of pathogen carried by the infected animals, the time between deposition of faecal matter on the roof and pathogens being flushed into the tank, the form of exposure (ingestion from drinking vs. exposure to droplets in the shower or toilet flushing), and the relative persistence of the different pathogens (Ahmed et al., 2011).

The detection and quantification of individual pathogenic bacteria in water quality evaluation

is labour intensive and not easy to perform in most cases. Hence, routine microbiological analyses of water are based on the detection of faecal indicator organisms, which share the same habitats (World Health Organisation, 2004). An ideal faecal indicator should be consistently present in faeces, unable to multiply outside the intestinal tract, be at least as resistant as the associated pathogens to environmental conditions and disinfection, have a strong association with the presence of pathogenic microorganisms, and be detectable by simple laboratory methods (Savichtcheva and Okabe, 2007; Alonso et al., 2010).

To determine the acceptability of RHRW for drinking, it is common practice to use drinking water guidelines. In most guidelines, this entails the non-detection of faecal indicator bacteria such as *E. coli* or enterococci (usually at numbers below 1 cfu/100ml water) whose presence are used to indicate potential faecal contamination of the water (World Health Organisation, 2004).

Historically, faecal indicator bacteria including *E. coli*, total coliforms and faecal coliforms and enterococci have been used in many countries as a monitoring tool for microbiological contamination of water and prediction of the presence of bacterial, viral and protozoan pathogens (Borrego et al., 1987; Ferguson et al., 1996). These microorganisms are mainly of faecal origin from higher mammals and birds, and their presence in water may indicate faecal pollution and possible association with enteric pathogens. In environmental waters, most faecal coliform strains are *E. coli*. However, *E. coli* presence in particular situations was reported to be definitively not associated with faecal pollution (Winfield and Groisman, 2003). It has also been shown that *E. coli* can persist and become naturalised in secondary, non-host habitats (Anderson et al., 2005).

The presence, persistence, and possible naturalisation of *E. coli* in aquatic habitats can impair the use faecal coliforms as a reliable indicator of recent faecal contamination in environmental waters (Ksoll et al., 2007; Lyautey et al., 2010). Considering these limitations, the use of enterococci is advisable and has been used over the years in the making of drinking water legislation. While faecal coliforms and *E. coli* are still widely used in environmental monitoring, enterococci are becoming a frequent target as they can be used to estimate health

risks in both recreational marine waters and freshwaters (Wright, 2008; Ryu et al., 2013). *Escherichia coli* is most useful in identifying recent contamination since it is not as environmentally long-lived as many pathogens (i. e. viruses and protozoa). However, additional complementary tests examining for the more robust enterococci and the spores of *Clostridium perfringens* can shed light on less recent faecal contamination (Ahmed et al., 2011).

Although *E. coli* is widely used to assess RHRW quality a number of recent studies have, reported that *E. coli* may be of limited use for comprehensive evaluation of harvested rainwater microbial quality. In these investigations a number of samples yielded culturable *Enterococcus* spp. but not *E. coli* (Spinks et al., 2006; Sazakli et al., 2007; Ahmed et al., 2008). It has consequently been suggested that HRW should be tested for multiple water quality indicator bacteria to obtain multiple lines of evidence on the occurrence of faecal contamination (Ahmed et al., 2010b, 2011).

Environmental waters are not a natural habitat for enterococci and their presence is considered the result of faecal pollution. Studies looking at enterococci diversity in environmental waters have identified most strains as *E. durans*, *E. faecalis*, *E. faecium* and *E. hirae*, *E. casseliflavus*, *E. mundtii* and less commonly, *E. avium*, *E. cecorum*, *E. columbae* and *E. gallinarum* (Lauková and Juris, 1996; Ferguson et al., 2005).

Enterococci are more resistant to stress and chlorination than *E. coli* and other coliform bacteria and have been used as the parameter of choice for faecal pollution and to complement *E. coli* in catchment assessment of tropical climates where *E. coli* is less appropriate because of the suspicion of multiplication (Vialle et al., 2011; Ahmed et al., 2012c). Enterococci are highly resistant to drying and thus may be valuable for RHRW monitoring. Pollution sources in RHRW include windblown dust and faecal droppings on the roof surface. These are subject to drying hence *E. coli* are likely to die while the more resilient enterococci will persist. Consequently, enterococci may serve as a better faecal pollution indicator and the pathogen of choice in bacterial source tracking (Ahmed et al., 2010b).

Enterococci are a diverse group of Gram-positive gastrointestinal colonisers with lifestyles ranging from intestinal symbionts, environmental persisters, to multidrug resistant nosocomial pathogens (Kühn et al., 2003). The *Enterococcus* genus includes more than 20 species, most of which are part of the intestinal flora of mammals, reptiles, birds, and other animals, while some species have been isolated from non-faecal sources (Devriese et al., 1993). Different groups of species predominate in different hosts, for example *E. faecalis* and *E. faecium* are dominant in the human digestive tract, whereas *E. cecorum*, *E. durans*, *E. faecalis*, *E. faecium* and *E. hirae* are dominant in poultry.

A number of researches have shown enterococci to be more prevalent in RHRW than *E. coli* and has been suggested to be a better indicator for assessing faecal contamination (Ahmed et al., 2011). However few studies have characterised *Enterococcus* spp. from RHRW (Ahmed et al., 2012b). In this study our aims were: (1) to investigate the prevalence of faecal indicators bacteria in harvested rainwater and alternative water sources used by rural household and in the various environmental settings that rainwater harvesting (RWH) is practiced in South Africa and (2) to evaluate the prevalence of *Enterococcus* spp. in RHRW and faecal droppings of pigeons as the most likely source of RHRW contamination.

## 4.2 Methodology

### 4.2.1 Sampling sites description and sample collection

Sampling areas were comprised of farm, urban residential and business, and rural settlement settings in Gauteng, Northwest and Eastern Cape Provinces. The sampling environments were divided into six sampling areas and included Johannesburg, (JHB, one area) and Pretoria (PTA, three areas) in Gauteng Province, Brits (BTS, one area) in the Northwest Province, and Port St Johns (PSJ, 1) in the Eastern Cape Province. Sites in Pretoria included the University of Pretoria Experimental Farm with three RWH tanks (PTA1, three tanks), the Plant Science Building, University of Pretoria Hatfield campus (PTA2, three tanks), a household in Sunnyside, Pretoria (PTA3, one tank). The Johannesburg sites included Thembisa township (JHB1, six tanks), and Weltevreden Park (JHB2, one tank) and the Harteesbeespoort, Ifafi site in Brits (BTS1, one tank). The Port St Johns (PSJ) sites included



15 RHRW tanks from fifteen households (PSJ1 to 15), four ground surface runoff harvested rainwater (GHRW) storage tanks, three rivers and one spring water source.

The PTA1 site represents a typical farm setting with a cattle pen within 50 meter of the rainwater tanks where masses of doves and pigeons feed on cattle feed. The house on which rainwater tanks were installed has overhanging mulberry trees on two sides of the roof, where various kinds of birds feed on mulberries. The roof catchment surface on these two sides feed water into two separate tanks. The third side of the roof is free of vegetation cover and feeds into a separate tank. Samples from this site were collected three times from beginning to mid rain season, whereas the rest of the samples from other sites were collected once off during mid-rain season (2012 to 2013).

The PTA2 site is located on the second floor of the Plant Science Building at the University of Pretoria main campus. Three tanks were installed and the site represents a typical urban business setting where there is minimal vegetation and bird interference. Water harvested at this site is used to irrigate flowers and experimental plants. The Thembisa site (JHB1) represents a typical urban township. Schools in this township were provided with RWH systems for irrigation of the school garden for the school feeding programme. The Weltevreden Park (JHB2) and Hartbeespoort (BTS1) sites have modern RWH systems with first flush diverters and a complex filtering system and were included for comparative purposes.

The PSJ site in the Eastern Cape Province, Port St Johns, is located in Luthengele village, which is situated in a mountainous area. The terrain in this village is such that municipal water supply services would be too expensive to implement. Hence, the people rely mainly on river and harvested rainwater (HRW). The community benefited from government projects where rainwater tanks were installed for potable and domestic food gardening purposes. In this village RHRW is stored in above-ground tanks and is used for potable purposes whereas GHRW is stored in underground tanks and used for domestic food gardening. Water from the local rivers and spring is used to supplement RHRW for potable purposes. To evaluate the influence of water handling practices, water collected from tanks

and stored prior to use in kitchens (hereafter referred to as kitchen water) was also collected (in two litre containers) from two households (PSJ1 and PSJ2) and for the rest of the households (PSJ3-15) samples were collected from RWH tanks only. In addition GHRW samples were also collected from four households that had installed these tanks on their homesteads.

Samples were collected in duplicates from the outlet taps located close to the base of the tanks, in sterilised two litre containers. Taps were wiped with 70% ethanol, and allowed to run for 30 to 60 seconds to flush out stagnant water from the taps before collecting water samples. Samples were transported to the laboratory and processed within 24 hours.

#### **4.2.2 Microbiological analysis of water samples**

Undiluted water samples were assayed directly for densities of faecal coliforms, *E. coli*, and enterococci with Colilert and Enterolert chromogenic substrate tests kits and Quantitray 2000 trays (Idexx, Westbrook, Maine) as per the manufacturer's instructions. Water samples (100mℓ) were poured into Quantitrays which were then sealed and incubated at 35°C (Colilert for faecal coliforms and *E. coli*) and 40.5°C (Enterolert for enterococci) for 24-28 hours. Following incubation, the Colilert Quantitray wells were read for yellow colour and fluorescence (faecal coliforms and *E. coli*, respectively), and Enterolert Quantitrays were read for fluorescence only (enterococci). A bench top ultra violet (UV) light (366 nm) was used to identify fluorescent wells. The manufacturer (Idexx, Westbrook, Maine) provided most probable number (MPN) of colony forming units (cfu) table, which was used to generate microbial density estimates based on the proportion of positive reactions in each tray.

#### **4.2.3 Faecal sample collection**

Thirty fresh faecal droppings were collected from pigeons that came to feed at the cattle feedlot of the University of Pretoria Experimental Farm. Indicator bacterial density in faecal material was measured by first diluting one gram of faeces in 9mℓ distilled water, vortexed and allowed to stand for five minutes to allow debris to settle. A 1mℓ sample of the supernatant was consequently extracted and serially diluted for microbial isolation and quantification. Densities of faecal coliforms and *E. coli*, and enterococci were determined

with Colilert-18 and Enterolert-18 chromogenic substrate tests kits and Quantitray 2000 trays (Idexx, Westbrook, Maine) as per the manufacturer's instructions. The diluted sample (1ml) was mixed with 99ml sterile water to meet the 100ml requirement for Quantitray 2000 trays for quantification.

#### **4.2.4 Recovery of isolates and presumptive identification**

Following incubation, the backing material of each Quantitray was disinfected by application of 70% ethanol with a sterile swab. After the residual ethanol evaporated, sterile razor blades were used to pierce the backing material of three fluorescence positive wells per tray and three trays were processed per water sample. One loop full of well content was streaked onto *Enterococcus* spp. selective agar (Merck, Johannesburg). Presumptive enterococci positive colonies were sub-cultured twice on nutrient agar (Merck, Johannesburg) and re-inoculated into 200µl of Enterolert chromogenic media (Idexx, Westbrook, Maine) in sterile 96 microwell plates and incubated for 18 hours to confirm fluorescence before matrix-assisted laser desorption ionisation (MALDI-TOF-MS) analysis.

#### **4.2.5 Matrix-assisted laser desorption ionisation time of flight mass spectroscopy identification and characterisation of bacterial isolates**

Bacterial strains were sub-cultured twice on nutrient agar (Merck, Johannesburg) before matrix-assisted laser desorption ionisation (MALDI-TOF-MS) analysis. After 24 to 48 hour cultivation of an isolate on nutrient agar, a single colony was transferred with a toothpick onto MALDI biotyper target plates in duplicate (Bruker Daltonics, Bremen, Germany). The preparation was overlaid with 1 µl of saturated solution of cyano-4-hydroxycinnamic acid in organic solution (50% acetonitrile, 2.5% trifluoroacetic acid), crystallised by air-drying at room temperature and directly screened (Bittar et al., 2009; Pinto et al., 2011). Mass spectra were generated with the Microflex LT mass spectrometer operated by the MALDI Biotyper automation control and recorded by Flex Control software (Bruker Daltonics, Bremen, Germany). Three hundred shots per sample spot were acquired using the recommended instrument settings for bacterial identification (linear positive mode, 60 Hz laser frequency, 20 kV acceleration voltage, 16.7 kV IS2 voltage, 170 ns extraction delay, and 2,000 to 20,137 m/z range). The peak lists generated was used for matches against the reference library directly using the integrated pattern-matching algorithm of the software. The whole process

from MALDI-TOF-MS measurement for identification was performed automatically without user intervention.

MALDI Biotyper 3.0 software (Bruker Daltonics, Germany) was used to analyse raw spectra of the bacterial isolates, with default settings. The software compares acquired sample spectra to reference spectra in the provided database and compiles a list of best matching database records. The degree of spectral pattern matching is expressed as a logarithmic identification score and interpreted according to the manufacturer's instructions. Results are expressed as log (score) values ranging from 0 to 3 levels. Scores  $\geq 2.300$  indicate species identification with a high level of confidence,  $\geq 2.000$  indicate species identification, 1.700 -1.999 indicate genus identification, and  $< 1.700$  no identification (Romanus et al., 2011).

#### **4.2.6 Statistical Analysis**

All statistical analysis were carried out using Statistica 10 (Stat soft, US). Data for the microbial concentrations were separated into areas JHB, PTA, BTS and PSJ. Water samples were divided into RHRW, GHRW and alternative sources (spring and river). To compare differences in microbial concentration graphical representations of the various indicator bacteria were used. The various enterococci species were grouped by their sources to compare their respective percentage prevalence.

### **4.3 Results**

#### **4.3.1 Samples from Gauteng Province (Johannesburg and Pretoria)**

Fourteen RWH tanks were sampled from Johannesburg and Pretoria sites. *Escherichia coli* were detected in nine tanks from five of the six sites: JHB1 (three tanks), PTA1 (three tanks), PTA2 (two tanks) and PTA3 (one tank). The concentrations of *E. coli* from two of the tanks from JHB1 were 3.1 and 15.3 cfu/100ml whereas the concentrations in the other tank was  $> 2149$  cfu/100ml. Although *E. coli* were detected in all the three tanks from site PTA1 the concentrations varied with the presence of overhanging trees as detailed below. At site PTA2 *E. coli* counts were 103.6 and 145 cfu/100ml in the two positive tanks. Faecal coliforms were detected in all the tanks at site PTA2 and the concentrations ranged from 10.9 to  $> 2419.6$

cfu/100mℓ with an average of 716.5 cfu/100mℓ. Enterococci were detected in the tanks from the sites JHB1 (3 tanks), JHB2 (1 tank) PTA2 (2 tanks) PTA3 (1 tank) and BTS1 (1 tank). The concentration of enterococci detected ranged from as low as 1.7 to >2419.6 cfu/100mℓ, with an average of 715.08 cfu/mℓ.

Samples were collected from two sites (JHB2 and BST1) that use modern RWH systems for all their domestic purposes including drinking, bathing and washing. The RWH systems were installed with first flush diverters and multiple filtration systems. Water samples were collected directly from the tank before filtration and at the point of use after filtration including water from the geyser. Water samples collected after filtration points from both sites tested negative for *E. coli*, faecal coliforms and enterococci. Water samples collected directly from the tank tested positive for enterococci at both sites (JHB2 and BST1), whereas, only the JHB2 site tested positive for faecal coliforms. The detected enterococci concentration observed from the JHB2 site was 31.8 cfu/100mℓ and that for the BST1 site was 7.5 cfu/100mℓ and the detected faecal coliform concentration were 8.6 cfu/100mℓ.

The three tanks from PTA1 site (PTA1 to 3) were installed on a house located close to a cattle feeding pen where masses of pigeons come to feed on cattle feed. At the house where the tanks were installed there were overhanging mulberry trees on two sides of the house from which two tanks (PTA1-1 and PTA1-2) received roof runoff. However, the third tank (PTA1-3) was installed on the side where there were no overhanging trees. Although *E. coli* were detected from all tanks during the three sample collections, the concentrations were less than 25 cfu/100mℓ except for the samples collected from tanks PTA1-1 and PTA1-2 during the third sampling event where the counts were >2419.6 cfu/100mℓ for both tanks. Concentrations detected from tank PTA3-1 during the three samplings were 4.13, 220 and 387.3 cfu/100mℓ for enterococci; 31.8, 574 and 688.4 cfu/100mℓ for faecal coliforms, whereas those for tanks PTA1-1 and PTA1-2 were >2419.6 cfu/100mℓ for all the samples, for both enterococci and faecal coliforms. It is evident that samples from tanks on the roof sides with over hanging trees had consistently higher concentrations of all the indicator bacteria compared to the samples from the tank on the side without overhanging trees.

#### 4.3.2 Prevalence of *Escherichia coli*, enterococci and *Pseudomonas* spp. in roof harvested rainwater samples from Luthengele village, Eastern Cape

The quality of water used by households in Port St Johns, Luthengele village (PSJ site) was evaluated based on *E. coli*, enterococci and *P. aeruginosa*. *Escherichia coli* were detected in seven of the fifteen roof harvested rainwater storage tanks and ranged from 1 to 8.6 cfu/100ml (Figure 4.1). Kitchen water was collected from two households (PSJ1 and PSJ2), and notable differences were observed when it was compared to the source tank water quality. At household PSJ1, enterococci and *P. aeruginosa* were not detected in tank water but in kitchen water at values of 1046.2 and 6.3 cfu/100ml respectively. At household PSJ2 *P. aeruginosa* was only detected in kitchen water (1986 cfu/100ml) (Figure 4.1). When all RHRW tanks were considered *P. aeruginosa* was detected in eight of the fifteen tanks and ranged from 1 to 120.1 cfu/100ml. Enterococci were detected in four of the fifteen tanks (PSJ5, 6, 12 and 14) and the highest concentrations were observed from tank PSJ12 (>2419 cfu/100ml).

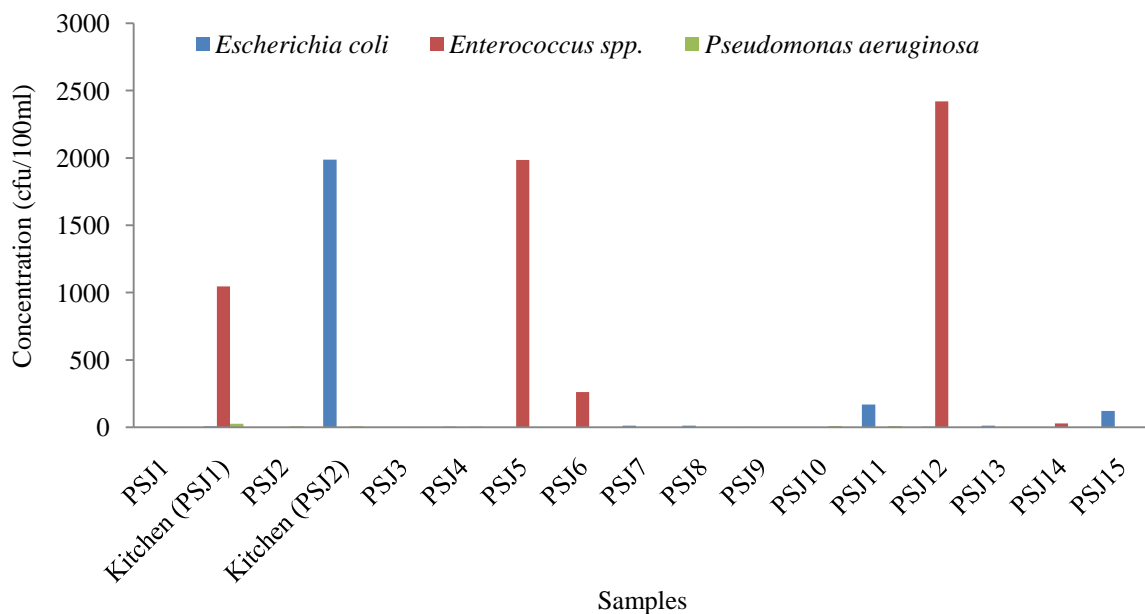


Figure 4. 1: Prevalence of bacterial species in tank and kitchen water samples. Kitchen water samples were collected from PSJ1 and PSJ2 tanks only

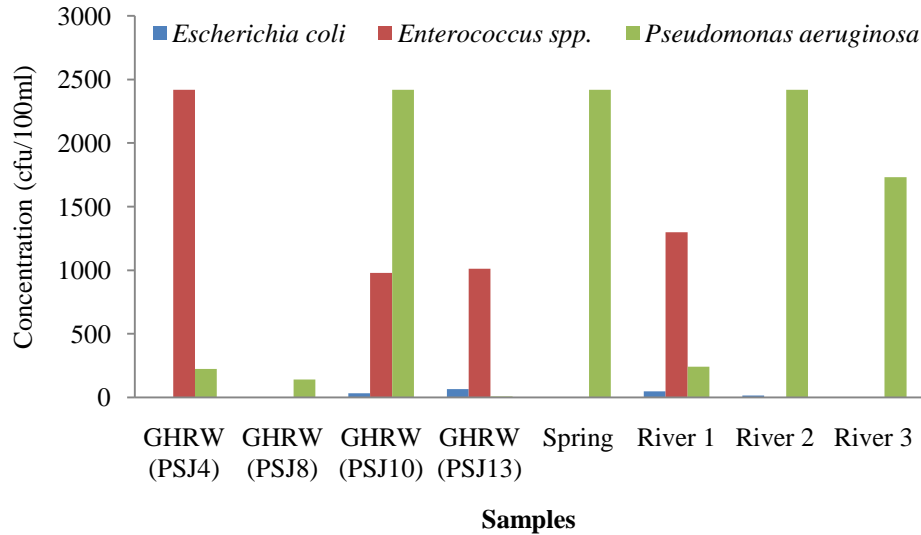


Figure 4. 2: Prevalence of bacterial species in ground harvested rainwater used for irrigation and river and spring water used for potable purposes in Luthengele village Eastern Cape Province, South Africa.

Three rivers and one spring water source used by households were sampled and analysed (Figure 4.2). Enterococci were only detected from one river (river 1) (1299 cfu/100ml). *Escherichia coli* were detected in all three rivers at 3 to 16 cfu/100ml. Only *P. aeruginosa* was detected in spring water at 2419 cfu/100ml. The three river water samples were positive for *P. aeruginosa* at concentrations ranging from 1732 to >2419 cfu/100ml. Rainwater harvested from ground surface runoff and stored in underground tanks was positive for *E. coli* in two of the four tanks at 33.6 cfu/100ml (PSJ10) and 64 cfu/100ml (PSJ13). Three tanks (PSJ4, 10 and 13) tested positive for enterococci (980 to >2419.6 cfu/100ml) and *P. aeruginosa* was detected in all the tanks, of which the highest concentrations were detected from site PSJ-10 at a concentration >2419 cfu/100ml and the least at PSJ13 (6.3 cfu/100ml). Interestingly PSJ10 had the highest concentration observed for *E. coli* (64.7 cfu/100ml) and enterococci (1011 cfu/100ml).

#### 4.3.3 Prevalence of enterococci in roof harvested rainwater and bird faecal samples

Considering the thirty fresh urban pigeon faecal samples collected from the cattle feeding lot close to site PTA1, nineteen and thirty of the samples tested positive for *E. coli* and enterococci, respectively. A total of 364 enterococci were isolated from thirty pigeon faecal

samples (156 isolates) and eleven RHRW tanks (208 isolates) from sites PTA1, PTA2 PTA3 and JHB1. MALDI-TOF-MS was used for the identification of the various enterococci species. In total seven enterococci species were identified (Figure 4.3). Four species *E. faecalis* (20.5%), *E. mundtii* (20.51%), *E. faecium* (23.1%) and *E. casseliflavus* (17.3%) were prevalent in faecal samples whereas *E. casseliflavus*(34.6%) and *E. mundtii*(33.2%) were prevalent in RHRW samples (Figure 4.3), although the concentrations of *E. faecalis* (18.7%) in RHRW were similar to those observed in faecal samples. The least abundant species were *E. durans* (2.5%) in both sample sources and *E. hadei* (4.5% in faeces and 2.9% in RHRW). Although *E. galinarium* was observed in low proportions in RHRW (0.96%), it constituted 11.5% of faecal isolates. Significant differences in the relative abundance of enterococci species were observed for *E. casseliflavus* and *E. hadei*. The prevalence of *E. casseliflavus* and *E. mundtii* in faecal samples was almost half the relative abundances detected in RHRW. *Enterococcus faecium* was three times more abundant in faecal samples (23.1%) than in RHRW samples (7.2%).

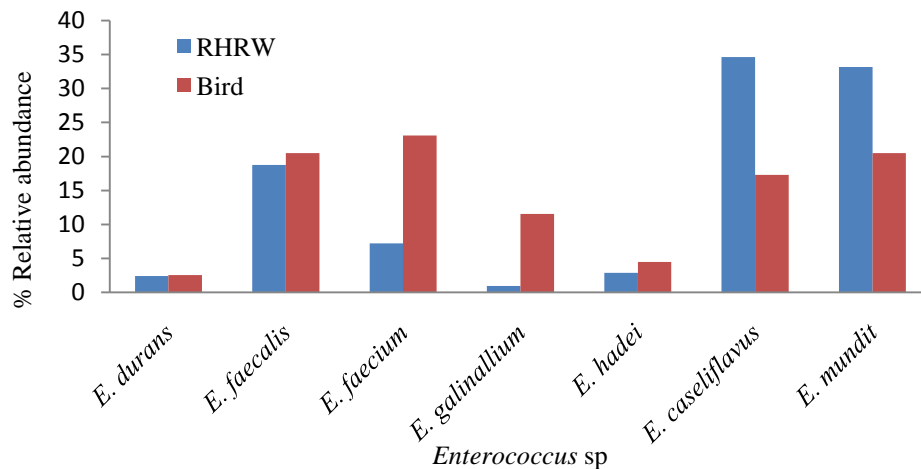


Figure 4. 3: Distribution of enterococci isolated from pigeon faeces and roof harvested rainwater from sites in Pretoria and Johannesburg.

#### 4.4 Discussion

The use of RHRW is currently increasing in South Africa, with both government and non-governmental organisations promoting the practice (Kahinda et al., 2007). Rainfall harvested from rooftops has been promoted for potable use, while rainfall collected from surface runoff has been promoted for domestic food gardening (Denison and Wotshela, 2009). In this study



we evaluated the prevalence of faecal coliforms, *E. coli*, enterococci and *P. aeruginosa* from urban, peri-urban, farm and rural settings. These sites were selected as part of a scoping study for a RWH project on the factors affecting the quality of RHRW and potential health risks (Water Research Commission, 2012).

#### **4.4.1 Factors affecting rain water quality**

The findings from the scoping study revealed significant variations in water quality between samples from various environmental settings. We observed the highest level of contamination from samples collected from a farm setting. The presence of faecal material on the ground surface, which can be blown by wind onto the roof surfaces, is one of the most critical factors affecting on farm contamination of roof harvested rainwater (Ahmed et al., 2012c). Our findings on the discrepancies between microbial counts in tanks installed at the same house but differing on the presence of overhanging trees revealed trees to be of great significance in the RHRW contamination. We observed that pigeons and doves that came to feed on the farm would rest in these trees and some even nested there. The presence of birds offered a constant supply of faecal matter on the roof surface, which is then washed into the tank during rain events. Overhanging trees provides shade to the roof surface, shielding it from direct sunlight, which has been shown to greatly enhance pathogen survival. In a research by Ahmed et al., (2010a) it was reported that bacteria could increase their period of survival by a factor of two under shading compared to direct sun light.

#### **4.4.2 Rainwater quality from different technological and environmental settings**

During our sampling visit we observed that the nature of the currently employed RWH systems varied from rudimentary to sophisticated systems. The most basic systems included a roof, gutter, and tank. A variety of makeshift containers from used drums, baths tubs, clay pots etc. are used by villagers as storage containers. In such situations, neither a first flush diverter nor a filtration system is employed. However, in modern RWH systems such as the ones from sites JHB2 and BTS1, first flush diverters were installed together with a filtration system. Consequently, these samples had the lowest level of contamination and the residents at the household where the tanks were installed appeared to be better informed about RWH system management and potential health risks compared to their rural counterparts. While the

use of modern rainwater systems would be ideal in terms of guaranteeing the water quality, the installation of such system would be too expensive for rural households who can hardly afford the cost of purchasing a RWH tank.

Households in Luthengele village in the Eastern Cape Province of South Africa use river water, RHRW and spring water for potable purposes. The water from these sources is generally used without prior treatment. The water from the rivers flows directly from the groundwater table and therefore should be relatively free of faecal contamination. The water is collected at some distance after the points where it comes off the ground and these collection points are shared by both people and livestock. Animal faeces are a common feature at these collection points. However, analysis of the microbiological quality of river water did not show it to be worse off than roof harvested rainwater in the observed microbiological counts. All the three river water collection points, tested positive for *P. aeruginosa*, while one and three of the collection points tested for enterococci and *E. coli*, respectively. Since these rivers are constantly flowing, the dilution effect play a significant role in the microbiological counts observed. The observation of higher concentration of *P. aeruginosa* may be explained by the fact that it is a common soil inhabitant (Pirnay et al., 2009).

While the observed microbiological counts may suggest the water to be relatively clean, river water is an open system that is subject to too many uncertain factors of contamination. Domestic animals drink water early in the morning and late in the afternoon at the same points where the people fetch water from. Since we collected our samples during mid-day from these points the observed microbial quantities may not give a true representation of the levels of contamination and potential risks. Hence, we cannot make a general statement concerning the microbial quality of these water sources. There is therefore a need to evaluate the variation in the microbial quality of the water along the day in order to come up with appropriate times at which the people can fetch water when levels of contamination will be at their lowest.

Water in South African rural communities is fetched from the source (tank, river, borehole,

etc.) using a variety of containers and stored prior to use. Although the water might be clean at the source, the process of fetching and storing prior to use exposes it to potential contamination. Comparison of the microbiological quality of RHRW sampled directly from the tanks to that which had been stored temporarily prior to use showed significant variations in their microbiological quality. While we did not detect any significant indicator bacteria quantities in the RHRW from both tanks tested, we detected high concentration of enterococci in kitchen water of one household and *P. aeruginosa* in the other. These findings suggest the containers used to store the water or the process of fetching it, to be responsible for the contamination detected. It would be advisable therefore that the containers be thoroughly cleaned. However the most common type of containers used to fetch water from RHRW tanks for storage in the kitchens prior to use are narrow mouthed 20 or 25L containers which are difficult to clean inside. Consequently bacteria may accumulate in the form of biofilm creating a perennial source of contamination (Camper et al., 1998).

Although *E. coli* was detected in two of the four ground harvested rainwater tanks tested, the counts observed for enterococci and *P. aeruginosa* in the same tanks were at least ten times higher. Enterococci and *P. aeruginosa* were detected in three of the four tanks and the observed concentrations of enterococci were relatively higher. The microbiological counts observed in this water especially with respect to *E. coli* may suggest the water not to be highly contaminated. However, consideration of alternative faecal indicator bacteria represented by enterococci suggests otherwise. The ground harvested rainwater is stored for relatively long periods of time in the tanks before use. It is therefore most likely that *E. coli* will die with time, while the more resilient enterococci and other pathogens persist (Ahmed et al., 2012a). Hence, there is need to evaluate the water quality by targeting specific pathogens and to evaluate the survival and persistence of the various pathogens in relation to the presence of the commonly used faecal indicators throughout the rainy season (Ahmed et al., 2010b).

#### **4.4.3 Roof harvested rainwater quality.**

The microbiological quality of roof harvested rainwater stored in tanks varied significantly from one household to another. While the water quality in some tanks were within the

drinking water standards, the microbiological quality of water from other tanks was so poor that treatment would be necessary before the water could be used for potable purposes. This variation supports the previous findings from a number of researched which have provided contradicting results on the quality of RHRW (Yaziz et al., 1989; Evans et al., 2006b; Meera and Ahammed, 2006; Ahmed et al., 2010b). While some researchers report that RHRW microbial quality is within the acceptable drinking water quality, others have reported the presence of indicator bacteria including pathogens and recommended pre-treatment before the water can be used for potable purposes (Ahmed et al., 2011). Our findings suggest that the quality of harvested rainwater stored in tanks is site specific and depends on the management practices implemented by the households as previously reported (Kus et al., 2010).

Faecal material on the roof surface has been cited as one of the major source of contamination in RHRW. Faeces of birds, insects, mammals, and reptiles that have access to the roof can potentially contain a wide array of pathogens (Ahmed et al., 2011, 2012a). During rain events, debris on the roof surface or in the gutter, including animal droppings can be washed by the roof runoff into the tank (Evans et al., 2007). Our results from the experimental farm are in agreement with this notion. The microbial quality of water from the roof side with overhanging mulberry trees was poor with high concentrations of indicator bacteria in contrast with the tank from the roof side without overhanging trees. The presence of high concentrations of indicator bacteria in samples, suggest the presence of a direct source of faecal contamination.

From our observations faecal contamination of RHRW appears to be limited to improperly designed systems, as well as systems that are not well maintained. Hence, guidance's on RHRW systems encourage good maintenance practices, including ensuring the cleanliness of the systems before rainfall events, especially roofs and gutters, which should be cleaned frequently, while the receiving tanks should be cleaned at least two times per year to improve the water quality (Cunliffe, 1998). The roof should be kept clear of overhanging trees, which may provide access to the roof by animals and birds. Indeed, the high numbers of bacteria in bird faecal samples indicate the need for good maintenance of roofs and gutters and

elimination of overhanging tree branches to minimise faecal contamination (Ahmed et al., 2012c).

The overall observation from the results of our study is that there are a number of critical factors responsible for the observed variations in roof harvested rainwater quality. It is clear that further data relating to the prevalence of microbial contamination including indicator bacteria and specific pathogens throughout the year and their persistence in rainwater tanks is needed. Considering the survival and persistence of bacterial species in RHRW, previous studies have suggested disinfection before use as potable water, especially for drinking (Ahmed et al., 2010b). It is however important that before treatment can be implemented, the levels of actual risk of contamination and infection should be established so that informed decisions can be made on management and mitigation practices that will be instituted in regulatory or guidance information.

#### **4.4.4 Indicator bacteria in roof harvested rainwater**

Heterotrophic plate counts, faecal coliforms, *E. coli*, and enterococci are the traditionally used bacteria of choice that have been adopted in most water quality regulations around the world (Moe et al., 1991; WHO, 1997), of which *E. coli* is the most used and recognised indicator bacteria (Edberg et al., 2000). However, our findings do not suggest *E. coli* to be a suitable indicator in RHRW systems, as has been previously reported (Ahmed et al., 2011). In this study, of the 15 RHRW samples from Luthengele village; five, eight and nine of the RHRW tanks samples tested positive for *E. coli*, enterococci and *P. aeruginosa*, whereas for the ground harvested rainwater and river water samples four, five and eight of the eight samples tested positive for *E. coli*, enterococci and *P. aeruginosa*, respectively. In most of the tanks from which *E. coli* were detected, the quantities were up to ten times lower than those for enterococci concentrations. Consequently, the suitability of *E. coli* as indicator bacteria in harvested rainwater systems has been questioned. Several studies have also reported higher prevalence of enterococci in RHRW tanks compared with *E. coli*. In a survey of forty-nine rainwater tanks for the presence of faecal coliforms, *E. coli*, and enterococci in Victoria, Australia, Spinks et al. (2006) reported that 33% were positive for *E. coli* and 73% positive for enterococci. In a study of twenty-seven rainwater tanks by Ahmed et al. (2011), 63% and

78% of the samples tested positive for *E. coli* and enterococci, respectively. In consideration of these findings and other published research findings, Ahmed et al. (2008) concluded that *E. coli* could not be detected in a number of the water samples that were positive for other indicators of potential faecal origin such as enterococci or *C. perfringens*. In testing of RHRW quality, multiple indicators should therefore be considered where possible, so as to obtain multiple sources of evidence on potential faecal contamination (Ahmed et al., 2011). On the basis of these findings it can be concluded that *E. coli* may be of limited use to assess the microbial quality of RHRW.

The higher prevalence of enterococci in RHRW may be a result of their better capacity to persist in water longer than *E. coli* (McFeters et al., 1974). A number of research papers in environmental water quality have also provided evidence that enterococci were a better faecal contamination indicator than *E. coli* (Savichtcheva and Okabe, 2007; Schets et al., 2010). However, there is need for further investigations into *E. coli* usefulness versus enterococci in RHRW studies. This should include the potential environmental sources of the particular indicator bacteria and their relative persistence in RHRW systems before any recommendations on the most suitable indicator bacteria can be made.

#### **4.4.5 Enterococci in roof harvested rainwater**

We used MALDI-TOF-MS to identify enterococci isolates from RHRW and bird faecal samples, so as to obtain information on their ecology, diversity and potential sources. The dominant enterococci identified in faecal samples were *E. faecalis* (20.5%), *E. mundtii* (20.51%), *E. faecium* (23.1%) and *E. casseliflavus* (17.3%), whereas *E. casseliflavus* (34.6%), *E. mundtii* (33.2%) and *E. faecalis* (18.7%) were dominant in RHRW samples from eleven tanks that tested positive. The presence of high numbers of *E. faecalis* (18.7%) and to a lesser extent *E. faecium* (7.2%) in five tank water samples from three of the six sampling sites, suggests the presence of faecal strains due to their high prevalence in warm blooded animals (Kühn et al., 2003). *Enterococcus faecalis* and *E. faecium* have consequently been suggested by a number of authors as potential candidates for microbial source tracking studies in environmental waters (Rodgers et al., 2003; Call and Plescia, 2008; Furtula et al., 2012).

The detection of *E. casseliflavus* as one of the dominant enterococci both in RHRW and bird faecal samples supports the previous finding which suggested that, although the species is considered to be mainly epiphytic (Mundt and Graham, 1968), its presence in the environment cannot be attributed exclusively to non-faecal sources. The presence of *E. Casseliflavus* in faeces could be attributed to it being incorporated into the microbiota of the digestive tract of birds after consumption (Layton et al., 2010).

Previous researches have attributed the high prevalence of *E. faecalis* in rainwater tank samples to *E. faecalis* being ubiquitous in nature (Ahmed et al., 2012a) and that it may not have limited host specificity as previously reported (Wheeler et al., 2002). Enterococci have been reported to be present in the faeces of non-human animals including birds similar to our findings (Kühn et al., 2003; Layton et al., 2010).

The presence of a variety of enterococci including *E. casseliflavus*, *E. faecalis*, *E. faecium*, *E. hirae* and *E. mundtii* in pigeon faecal samples has previously been reported (Radimersky et al., 2010). The species distribution among enterococci in pigeons in the study areas has not been investigated previously. In this study, we speciated enterococci isolated from thirty pigeon faecal samples and the species. *E. casseliflavus*, *E. faecalis*, *E. faecium*, *E. hirae* and *E. mundtii*, *E. galinarium*, *E. durans* and *E. hadei* were detected. The prevalence of *E. faecalis* (20.5%), *E. mundtii* (20.5%), *E. faecium* (23.1%) and *E. casseliflavus* (17.3%) in bird faecal samples suggests that the birds may have contributed the enterococci detected in rainwater tanks. It is also possible that other sources of enterococci not tested in this study, including dust, vegetation and animals such as, lizards, rats or frogs may have contributed to the enterococci detected in RHRW tanks (Evans et al., 2006; Ahmed et al., 2012b).

The presence of *E. mundtii* and *E. casseliflavus* in ten of the eleven tanks sampled is not surprising, considering that these species have been reported to be associated with soil, plants and non-human animal hosts (Pinto et al., 1999). In this study, *E. mundtii* and *E. casseliflavus* constituted 67.8% in RHRW isolates and 37.8% in birds, supporting the notion that although birds are the most likely source of contamination, they are not the sole source of these

bacteria. These findings show the significance of enterococci speciation in its use as indicator bacteria of faecal contamination. Furthermore, the lack of one dominant enterococci species in both faecal and RHRW samples suggests that no single enterococci is a reliable indicator of the host faecal source. However, strain diversity characterisation of enterococci have been suggested to provide supporting evidence for bacterial source identification whether it be faecal, vegetation or soil (Layton et al., 2010)

#### **4.5 Conclusions and recommendations**

The contamination of HRW appears to be strongly influenced by the environmental settings especially the presence of a faecal source in the form of animal housing. While little can be done to the presence of a faecal source around rural households practicing animal husbandry, appropriate RWH system maintenance should be in place to lessen the levels of contamination. This should include system cleanliness by frequent cleaning especially of roofs and gutters before rainfall events. Roofs should be kept clear of overhanging trees, which may provide access to the roof by animals especially birds.

It is however evident that further information relating to the actual levels of risk as defined by the occurrence of pathogens in faeces of domestic animals should be evaluated. This should include the evaluation of the survival and persistence of bacterial and protozoan pathogens in rainwater tanks. Although a number of researches have suggested disinfection of RHRW before use, households in the rural communities in this study do not practice any form of water treatment. The quality of water used by these households could be greatly improved by implementing effective point of use treatment procedures, which includes filtration and disinfection. Such procedures should take into consideration the poor financial status of the rural communities for the solutions to be sustainable.

In this study, we focused on indicator bacteria that however do not give a clear indication of the actual level of risk from the water sources. It is therefore necessary that further studies focusing on the prevalence of specific pathogens including the characterisation of specific pathogenic factors be undertaken. This research formed a scoping study on the microbiological quality of RHRW and alternative water sources in rural communities. The



detailed research (currently underway) will include the above stated needs in microbial quality evaluation of the water sources.

#### 4.6 References

- Ahmed, W., A. Goonetilleke, and T. Gardner. 2010a. Implications of faecal indicator bacteria for the microbiological assessment of roof harvested rainwater quality in southeast Queensland, Australia. *Canadian Journal of Microbiology* 56: 471–479.
- Ahmed, W., A. Vieritz, A. Goonetilleke, and T. Gardner. 2010b. Health risk from potable and non-potable uses of roof harvested rainwater using quantitative microbial risk assessment in Australia. *Applied and Environmental Microbiology* 76: 7382–7391.
- Ahmed, W., F. Huygens, A. Goonetilleke, and T. Gardner. 2008. Real-Time PCR detection of pathogenic microorganisms in roof harvested rainwater in Southeast Queensland, Australia. American Society for Microbiology (ASM). *Applied and Environmental Microbiology* 74: 5490–5496.
- Ahmed, W., J. P. S. Sidhu, and S. Toze. 2012b. Speciation and frequency of virulence genes of *Enterococcus* spp. isolated from rainwater tank samples in Southeast Queensland, Australia. *Environmental Science Technology* 46: 6843–6850.
- Ahmed, W., J. P. S. Sidhu, S. Toze, and L. Hodgers. 2012c. Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland. *Applied and Environmental Microbiology* 5: 219–226.
- Ahmed, W., L. Hodgers, J. Sidhu, T. Gardner, K. Richardson, A. Palmer, and S. Toze. 2012a. Health risk assessment of roof- captured rainwater. Urban Water Security Research Alliance Technical Report No. 77.
- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Alonso, L., J. Rodriguez-manzano, R. Girones, M. A. Ferru, A. Hundesa, A. Carratala, B. Calgua, and A. D. A. Corre. 2010. Molecular detection of pathogens in water: The pros and cons of molecular techniques. *Water Research* 44: 4325-4339.

- Anderson, K. L., J. E. Whitlock, and V. J. Harwood. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied and Environmental Microbiology* 71: 3041–3048.
- Bittar, F., Z. Ouchenane, F. Smati, D. Raoult, and J. -M. Rolain. 2009. MALDI-TOF-MS for rapid detection of staphylococcal Panton-Valentine leukocidin. *International Journal of Antimicrobial Agents* 34: 467–470.
- Borrego, J. J., M. A. Moriñigo, A. de Vicente, R. Córna, and P. Romero. 1987. Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms. *Water Research* 21: 1473–1480.
- Call, D. R., and P. Plescia. 2008. Identifying sources of fecal pollution in the Colville River using library-independent genetic markers. *Northwest Science* 82: 120–127.
- Camper, A. M., B. Ellis, P. Butterfield, and C. Abernathy. 1998. Development and structure of drinking water biofilms and techniques for their study. *Journal of Applied Microbiology* 85: 1S–12S.
- Cox, P., M. Griffith, M. Angles, D. Deere, and C. Ferguson. 2005. Concentrations of pathogens and indicators in animal faeces in the Sydney watershed. *Applied and Environmental Microbiology* 71: 5929–5934.
- Cunliffe, D. A. 1998. Guidance on the use of rainwater tanks. *National Environmental Health Forum Monographs. Water Series 5210*: 1–29.
- Curtis, V., S. Cairncross, and R. Yonli. 2000. Review: Domestic hygiene and diarrhoea pinpointing the problem. *Tropical Medicine and International Health* 5: 22–32.
- Denison, J., and L. Wotshela. 2009. Indigenous water harvesting and conservation practices: Historical context, cases and implications. WRC Report No. TT 392/09. Water Research Commission (WRC). Pretoria.
- Devriese, L. A., B. Pot, and M. D. Collins. 1993. Phenotypic identification of the genus *Enterococcus* and differentiation of phylogenetically distinct enterococcal species and species groups. *Journal of Applied Bacteriology* 75: 399–408.
- Dillaha, T. A., and W. J. Zolan. 1985. Rainwater catchment water quality in Micronesia. *Water Research* 19: 741–746.

- Edberg, S. C., E. W. Rice, R. J. Karlin, and M. J. Allen. 2000. *Escherichia coli*: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology* 88: 106S–116S.
- Evans, C. A., P. J. Coombes, and R. H. Dunstan. 2006. Wind, rain and bacteria: The effect of weather on the microbial composition of roof harvested rainwater. *Water Research* 40: 37–44.
- Evans, C. A., P. J. Coombes, R. H. Dunstan, and T. Harrison. 2007. Identifying the major influences on the microbial composition of roof harvested rainwater and the implications for water quality. *Water Science and Technology* 55: 245–253.
- Ferguson, C. M., B. G. Coote, N. J. Ashbolt, and I. M. Stevenson. 1996. Relationships between indicators, pathogens and water quality in an estuarine system. *Water Research* 30: 2045–2054.
- Ferguson, D. M., D. F. Moore, M. A. Getrich, and M. H. Zhouandai. 2005. Enumeration and speciation of enterococci found in marine and intertidal sediments and coastal water in southern California. *Journal of Applied Microbiology* 99: 598–608.
- Furtula, V., Jackson, C. R., Osman, R., & Chambers, P. A. 2012. Use of *Enterococcus* spp. BST and Sterols for Poultry Pollution Source Tracking in Surface and Groundwater, *Environmental Health - Emerging Issues and Practice*, Prof. Jacques Oosthuizen (Ed.), ISBN: 978-953-307-854-0, In Tech, DOI: 10.5772/30191. Available from: <http://www.intechopen.com/books/environmental-health-emerging-issues-and-practice/use-of-enterococcus-bst-and-sterols-as-indicators-for-poultry-pollution-source-tracking-in-surface-a>
- Heyworth, J. S., G. Glonek, E. J. Maynard, P. A. Baghurst, and J. Finlay-Jones. 2006. Consumption of untreated tank rainwater and gastroenteritis among young children in South Australia. *International Journal of Epidemiology* 35: 1051–1058.
- Islam, M. A., H. Sakakibara, M. R. Karim, M. Sekine, and Z. H. Mahmud. 2011. Bacteriological assessment of drinking water supply options in coastal areas of Bangladesh. *Journal of Water and Health* 9: 415–428.
- Jagals, P., W. O. K. Grabow, and J. C. De Villiers. 1995. Evaluation of indicators for assessment of human and animal faecal pollution of surface runoff. *Water Science and Technology* 31: 235–241.

- Kahinda, J. M., A. E. Taigbenu, and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 32: 1050–1057.
- Kahinda, J. M., and A. E. Taigbenu. 2011. Rainwater harvesting in South Africa: Challenges and opportunities. *Physics and Chemistry of the Earth Parts ABC* 36: 968–976.
- Ksoll, W. B., S. Ishii, M. J. Sadowsky, and R. E. Hicks. 2007. Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. *Applied and Environmental Microbiology* 73: 3771–3778.
- Kühn, I., A. Iversen, L. G. Burman, B. Olsson-Liljequist, A. Franklin, M. Finn, F. Aarestrup, A. M. Seyfarth, A. R. Blanch, X. Vilanova, H. Taylor, J. Caplin, M. A. Moreno, L. Dominguez, I. A. Herrero, R. Möllby. 2003. Comparison of enterococcal populations in animals, humans, and the environment: A European study. *International Journal of Food Microbiology* 88: 133–145.
- Kus, B., J. Kandasamy, S. Vigneswaran, and H. K. Shon. 2010. Water quality characterisation of rainwater in tanks at different times and locations. *Water Science and Technology* 61: 429–439.
- Lauková, A., and P. Juris. 1996. Distribution and characterisation of *Enterococcus* species in municipal sewages. *Microbios* 89: 73–80.
- Layton, B. A., S. P. Walters, L. H. Lam, and A. B. Boehm. 2010. *Enterococcus* species distribution among human and animal hosts using multiplex PCR. *Journal of Applied Microbiology* 109: 539–547.
- Lyautey, E., Z. Lu, D. R. Lapen, G. Wilkes, A. Scott, T. Berkers, T. A. Edge, and E. Topp. 2010. Distribution and diversity of *Escherichia coli* populations in the South Nation River drainage basin, Eastern Ontario, Canada. *Applied and Environmental Microbiology* 76: 1486–1496.
- Lye, D. J. 2002. Health risks associated with consumption of untreated water from household roof catchment systems. *Journal of the American Water Resources Association* 38: 1301–1306.

- McFeters, G. A., G. K. Bissonnette, J. J. Jezeski, C. A. Thomson, and D. G. Stuart. 1974. Comparative survival of indicator bacteria and enteric pathogens in well water. *Applied Microbiology* 27: 823–829.
- Meera, V., and M. M. Ahammed. 2006. Water quality of rooftop rainwater harvesting systems: A review. *Journal of Water Supply Research and Technology AQUA* 55: 257–268.
- Moe, C. L., M. D. Sobsey, G. P. Samsa, and V. Mesolo. 1991. Bacterial indicators of risk of diarrhoeal disease from drinking water in the Philippines. *World Health Organisation. Bulletin of the World Health Organisation* 69: 305-317.
- Mpogui, A., and A. C. Mpogui. 2012. Consumption of rainwater and diarrheal disease in children under five in the Dominican Republic from 2002 to 2007. Msc Thesis, Georgia State University, [http://scholarworks.gsu.edu/iph\\_theses/232](http://scholarworks.gsu.edu/iph_theses/232)
- Mundt, J. O., and W. F. Graham. 1968. *Streptococcus faecium* var. *casseliflavus*, *Journal of Bacteriology* 95: 2005–2009.
- Pinto, B., R. Pierotti, G. Canale, and D. Reali. 1999. Characterisation of “faecal streptococci” as indicators of faecal pollution and distribution in the environment. *Letters in Applied Microbiology* 29: 258–263.
- Pinto, L., P. Poeta, H. Radhouani, C. Coelho, C. Carvalho, J. Rodrigues, C. Torres, R. Vitorino, P. Domingues, and G. Igrejas. 2011. Proteomic study in an *Escherichia coli* strain from seagulls of the Berlengas Natural Reserve of Portugal. *Journal of Integrated Omics* 1: 36–41.
- Pirnay, J., F. Bilocq, B. Pot, P. Cornelis, M. Zizi, J. Van Eldere, M. Vaneechoutte, S. Jennes, T. Pitt, and D. De Vos. 2009. *Pseudomonas aeruginosa* population structure revisited. *PLoS One* 4: e7740.
- Radimersky, T., P. Frolkova, D. Janoszowska, M. Dolejska, P. Svec, E. Roubalova, P. Cikova, A. Cizek, and I. Literak. 2010. Antibiotic resistance in faecal bacteria (*Escherichia coli*, *Enterococcus* spp.) in feral pigeons. *Journal of Applied Microbiology* 109: 1687–1695.

- Rodgers, K., P. G. Hartel, R. L. Kuntz, D. G. Godfrey, and W. I. Segars. 2003. Field-testing *Enterococcus faecalis* for bacterial source tracking. In The Proceedings of the 2003 Georgia Water Resources Conference. University of Georgia, Athens, Georgia, USA.
- Romanus, I. I., A. T. Eze, O. A. Egwu, A. N. Felicitas, and N. A. Chidiebube. 2011. Comparison of matrix-assisted laser desorption ionisation time of flight mass spectrometry (maldi-tofms) with conventional culture and biochemical method of bacteria identification to species level. *Journal of Medical Laboratory and Diagnosis* 2: 1–4.
- Ryu, H., M. Henson, M. Elk, C. Toledo-Hernandez, J. Griffith, D. Blackwood, R. Noble, M. Gourmelon, S. Glassmeyere and S. J. Domingo. 2013. Development of quantitative PCR assays targeting 16S rRNA genes of *Enterococcus* spp. and their application to the identification of *Enterococcus* species in environmental samples. *Applied and Environmental Microbiology* 79: 196–204.
- Savichtcheva, O., and S. Okabe. 2006. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. *Water Research* 40: 2463–2476.
- Sazakli, E., A. Alexopoulos, and M. Leotsinidis. 2007. Rainwater harvesting, quality assessment and utilisation in Kefalonia Island, Greece. *Water Research* 41: 2039–2047.
- Schets, F. M., R. Italiaander, H. H. J. L. Van Den Berg, and A. M. De RodaHusman. 2010. Rainwater harvesting: quality assessment and utilisation in The Netherlands. *Journal of Water and Health* 8: 224–235.
- Sinton, L. W., R. K. Finlay, and D. J. Hannah. 1998. Distinguishing human from animal faecal contamination in water: A review. *New Zealand Journal of Marine and Freshwater Research* 32: 323–348.
- Spinks, J., S. Phillips, P. Robinson, and P. Buynder. 2006. Bushfires and tank rainwater quality: A cause for concern? *Journal of Water and Health* 4: 21–28.
- Vialle, C., C. Sablayrolles, M. Lovera, S. Jacob, M. -C. Huau, and M. Montrejaud-Vignoles. 2011. Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. *Water Research* 45: 3765–3775.

- Water Research Commission (WRC). 2012. Evaluation of the risks associated with the use of rain-water harvested from rooftops, for domestic use and, homestead food gardens and groundwater for domestic use and livestock watering. WRC Project No 2175/4. Pretoria, South Africa.
- Wheeler, A. L., P. G. Hartel, D. G. Godfrey, J. L. Hill, and W. I. Segars. 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. *Journal of Environmental Quality* 31: 1286–1293.
- Winfield, M. D., and E. A. Groisman. 2003. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Applied and Environmental Microbiology* 69: 3687–3694.
- World Health Organisation 1997. *Guidelines for Drinking-Water Quality: Surveillance and Control of Community Supplies Vol 3*. World Health Organization. Geneva.
- World Health Organisation. 2004. *Guidelines for drinking water quality, vol 1 Recommendations*. World Health Organisation, Geneva.
- Wright, M. E. 2008. Evaluation of enterococci, an indicator microbe, and the sources that impact the water quality at a subtropical non-point source recreational beach. M. Sc Thesis, University of Miami.
- Yaziz, M. I., H. Gunting, N. Sapari, and A. W. Ghazali. 1989. Variations in rainwater quality from roof catchments. *Water Research* 23: 761–765.

## CHAPTER FIVE

### MALDI-TOF-MS CHARACTERISATION OF *ESCHERICHIA COLI* ISOLATED FROM ROOF HARVESTED RAINWATER

#### Abstract

The development of efficient techniques to identify and distinguish between *Escherichia coli* from different sources in aquatic environments is essential to improve the surveillance of faecal pollution indicators. Optimised data analysis methods can further contribute to develop strategies to identify sources of faecal contamination and to implement appropriate management practices to minimise gastrointestinal disease transmission. Currently available bacterial source tracking tools are often technically demanding, time consuming, and have limited accuracy in grouping isolates according to their respective sources. We evaluated the use of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) for the characterisation of environmental isolates of *E. coli*. Our main objectives were to optimise spectra processing and analysis protocols for obtaining unique MALDI-TOF-MS strains. After data pre-processing step, the resulting high quality data set was used for unsupervised hierarchical clustering using principle component analysis data (PCA) at 95% explained variance. To evaluate the lowest distance at which a single strain would cluster, 12 replicate spectra were included. These clustered at a similarity distance (SD) of 7.34 and the cut-off for unique strains was set at similarity distance of 10 and a total of thirty-one clusters were selected. The spectra of isolates in 31 clusters, representing unique strains were put in respective strain groups. Principal component analysis was used to evaluate how strain groups were distributed in a three dimensional (3D) plane and to establish a mini-optimised profile for *E. coli* strain classification. Support vector machine (SVM) was used to evaluate the uniqueness of the strain groups and to generate classification algorithms that would distinguish between the different stain groups. The support vector machines technique resulted in overall accuracies between 94% and 98% in distinguishing thirty-one strain groups.

Key Words: MALDI-TOF-MS, *E. coli*, support vector machine, bacterial source tracking



## 5.1 Introduction

Harvested rainwater (HRW) can potentially be contaminated by faecal matter deposited on catchment surfaces and is washed down into the tank during rainfall events. While efforts are ongoing to minimise the contamination of HRW, successful mitigation is limited when the source of the pollution is unknown (Ahmed et al., 2011a). The most cost-effective measure to reduce faecal pollution in HRW is to identify and mitigate the pollution at its source. This is best done by microbial source tracking (MST), which seeks to determine the origin of faecal material in water. It is therefore important to develop tools that can effectively detect and characterise faecal pathogens (Meays et al., 2004). Most MST methods are founded on the principle that, *E. coli* strains in the gastrointestinal tracts of different animals are of distinct subtypes (Ma et al., 2011). Although *E. coli* strains from a host source can display a wide range of diversity, strains that are highly similar have been reported to be from the same host source group (Carlos et al., 2012). Hence, matching the subtypes of *E. coli* identified in polluted HRW to those isolated from known sources would hypothetically allow the identification of the source of pollution.

Although most of the *E. coli* identified in roof harvested rainwater (RHRW) are non-pathogenic (commensal), their occurrence warns against the simultaneous occurrence of pathogenic bacteria (Ahmed et al., 2011b). Standard microbiological methods used for the detection of *E. coli* provide no information concerning host source. Identifying the source of faecal pollution is of high priority in order to better understand potential health risks and to mitigate the source of pollution (Meays et al., 2004). Faecal pollution can contain human pathogens such as *Salmonella* spp., *Shigella* spp., *E. coli*, enteroviruses, including hepatitis A virus, *Cryptosporidium parvum* and *E. coli* 0157:H7 (Ahmed et al., 2011a).

Several methods of bacterial source tracking (BST) methods have been developed in which sources of microbial contamination are identified according to differences in the characteristics of bacteria from different animal faeces (Meays et al., 2004). Most of the methods are library-dependent and use either phenotypic or genotypic profiles of the bacteria. Genotypic methods consider the host-specific genetic differences of indicator organisms, while phenotypic methods employ host-specific biochemical properties (Carlos et al., 2012).

Phenotypic methods include antibiotic resistance profiling, carbon source utilisation tests and Matrix-assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF-MS) profiling (Harwood et al., 2003; Giebel et al., 2008). Genotypic methods include ribotyping, pulse-field gel electrophoresis (PFGE), randomly amplified polymorphic DNA (RAPD), denaturing gradient gel electrophoresis (DGGE), repetitive DNA sequences (Rep PCR), terminal restriction fragment length polymorphism analysis (T-RFLP), host-specific 16S rDNA toxin biomarkers (Simpson et al., 2002; Field et al., 2003; Myoda et al., 2003). While many of these methods are promising, most are complex, costly, and often require days to complete. Reducing the time required for many of these BST tools is of crucial importance, since delays in determining faecal contamination sources increases the risk for human exposure to pathogens present in contaminated water.

Several of the existing BST tools cannot simultaneously identify and assign unknown isolates of the indicator organism to their respective sources. Most of the tools can distinguish between isolates of the same species but cannot differentiate between different species. For instance, when Stoeckel et al. (2004) classified *E. coli* isolates by sources, five out of seven (antibiotic resistance patterning, ribotyping using the EcoR1 restriction enzyme, PFGE, Box-PCR, and Rep PCR) BST tools met the minimum expectations for correct classification, with average correct source assignments ranging from 22% to 90%. Several factors contribute to such low accuracy levels including variation in data collection and analysis, variability in isolates collected at different times, sharing of *E. coli* subtypes between different sources, and fingerprint library size. Each of these problems can be associated with many of the BST tools currently used (Giebel et al., 2008).

The development of more accurate and rapid methods of characterising sources of bacterial contamination of environmental waters is desired considering the labour, time-intensive and often inaccurate nature of many existing BST tools (Giebel et al., 2008). MALDI-TOF-MS characterisation of bacteria is one promising method (Lay, 2001). In recent years, several reports have shown the feasibility of using MALDI-TOF-MS to identify microorganisms (Siegrist et al., 2007; Bittar et al., 2009; Dieckmann et al., 2010; Decristophoris et al., 2011; Fox et al., 2011). In whole-cell MALDI-TOF-MS, microbial fingerprints spectra of protein biomarkers are obtained from intact whole cells without biomarker prefractionation, digestion, separation, or cleanup. The procedure requires minimal amounts of biological material, is very fast and suitable for high-

throughput routine analysis. Hence, it has great potential for applications in environmental monitoring. Typically, house-keeping proteins that are highly expressed and conserved in bacteria as ribosomal or nucleic acid-binding proteins are used as biomarkers. Hence, the method can be universally applied and has been used for the classification/identification of *Salmonella* at the species and subspecies level (Dieckmann et al., 2008), detection of plasmid insertion in *E. coli*, differentiation between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*, detection of peptaibol formation by the genus *Trichoderma/Hypocrea* (Neuhof et al., 2007), speciation of staphylococci and their discrimination from micrococci (Fox et al., 2010) and discrimination between wild-type and ampicillin-resistant *E. coli* strains.

Although many recent studies have reported on the identification and characterisation of bacteria using MALDI-TOF-MS, reports on the application of this tool to BST of faecal pollution indicator bacteria are few (Giebel et al., 2008). MALDI-TOF-MS has been used to characterise and group a variety of mammalian and avian strains of *E. coli* by their sources (Siegrist et al., 2007) and for identification of clinically relevant bacteria (Prod'hom et al., 2010; Carbonnelle et al., 2011; Benagli et al., 2011).

MALDI-TOF-MS generated bacterial spectra have common characteristics, including sharp signal peaks with the associated baseline noise level and heteroscedasticity. These factors make direct analysis of raw mass spectra difficult, hence specific data analysis procedures are needed (Bruyne et al., 2011). A number of data analysis strategies can be applied. Although simple data analysis techniques including calculation of similarity coefficients are typically performed, they are less effective in distinguishing bacterial species at strain level (Hsieh et al., 2008). Hence, the more complicated techniques such as machine learning methods would be useful for strain level differentiation. The application of machine learning techniques for microbial species and identification and differentiation purposes is still limited. Currently, artificial neural networks (ANNs) dominate analysis of such data. However, they are slow to execute (Dieckmann et al., 2010; Bruyne et al., 2011; Tong et al., 2011).

In this paper, we apply the popular machine learning techniques, SVMs to differentiate between

*E. coli* strains of which only a small number of studies on MALDI-TOF-MS data for the identification and differentiation bacteria have been done (Harz et al., 2005; Hsieh et al., 2008; Bruyne et al., 2011). The majority of machine learning research papers to date have focused on the identification of particular species and do not include strain differentiation (Hsieh et al., 2008; Bruyne et al., 2011). This is, however, surprising as machine learning techniques could contribute significantly to the field of BST which relies on strain differentiation. Machine learning techniques can learn from data patterns and maximally exploiting embedded information. The learning process by SVM is typically fast and is important optimisation of the strain differentiation process. As a result, SVM can provide solutions in situation where standard analysis cannot, as they can easily handle multi-dimensional data (Hsieh et al., 2008; Bruyne et al., 2011).

In this study, we examined the potential of MALDI-TOF-MS-based bacterial fingerprinting to distinguish between *E. coli* environmental strains. Our specific objectives were to (i) evaluate whole-cell MALDI-TOF mass spectra accuracy in identification of *E. coli* environmental isolates (ii) determine whether the similarity of MALDI-TOF-MS profiles of isolates from the same source are adequately similar to allow grouping according to their respective sources, and (iii) to evaluate SVM data analysis methods ability to group isolates according to their respective sources.

## 5.2 Methodology

A total of 395 *E. coli* isolates from roof harvested rainwater (155) and pigeon faeces (136) were compared with those from sewerage (12), humans (53), bovine milk (5) and river water (18). Isolates from roof harvested rainwater and birds were collected from nineteen pigeon faecal samples and eleven roof harvested rainwater tanks that had tested positive for *E. coli*. These isolates were obtained during our survey of the microbial quality in RHRW stored in tanks and the prevalence of indicator bacteria in pigeon faeces as the most likely sources of contamination (Chapter 4). This included *E. coli* from RHRW that were isolated from two sites in Johannesburg (JHB1 and JHB2) and three sites in Pretoria (PTA1, PTA2 and PTA3). PTA1 is at the Pretoria included the University of Pretoria Experimental Farm and PTA2 is at the Plant Science Building, University of Pretoria Hatfield campus. The third site (PTA3) is a household in Sunnyside suburb

in Pretoria. The detailed description of the sites is provided in section 4.2.1. The rest of the isolates were obtained from the MALDI-TOF-MS laboratory in the Department of Microbiology and Plant Pathology at the University of Pretoria.

### **5.2.1 The Matrix-assisted laser desorption ionisation time of flight mass spectroscopy and characterisation of bacterial isolates were described in Chapter 4**

### **5.2.2 Polymerase chain reaction for detection of UidA gene in *Escherichia coli***

Isolates were confirmed to be *E. coli* by the polymerase chain reaction (PCR) detect for the presence of the UidA gene, which codes for the  $\beta$ -D-glucuronidase enzyme. A 147 bp coding region of the *E. coli* uidA gene was amplified using the 20 and 21-mer primers UAL-754 (5'-AAAACGGCAAGAAAA-AGCAG-3') and UAR-900 (5'-ACGCGTGGTTACAGTCTTGCG-3') (Bej et al., 1991). These primers were synthesised by Inqaba Biotech (Pretoria, South Africa). An ICycler thermal cycler (Bio-Rad, UK) was used to amplify the DNA. In both PCRs the final reaction volume of 25 $\mu$ l consisted of 12.5 $\mu$ l double strength PCR master mix (0.05U/ $\mu$ l *Taq* DNA Polymerase in reaction buffer, 0.4mM of each dNTP (dATP, dCTP, dGTP, dTTP), 4mM MgCl<sub>2</sub>; (Fermentas Life Science, US), PCR-grade water (Fermentas Life Science, US), 50ng sample DNA and 25pmole of each primer. Amplification was performed with a thermal cycler programmed for 1 cycle of 2 min at 94°C; 25 cycles of 1min at 94°C, 1 min at 58°C, 2 min at 72°C and 1 cycle of 5min at 72°C. PCR products were electrophoresed on a 1.5% agarose gel, stained with GR Green (Fermentas Life Science, USA), and visualised using Gel Documentation System (Bio-Rad Gel Doc EZ Imager, USA).

### **5.2.3 Data evaluation**

Spectra were imported into ClinProTools software (ver. 3.0; Bruker Daltonics) for post-processing. Spectra were processed with the following settings: peak definition (signal to noise ratio >3), integration (end-point level), mass recalibration (maximal peak shift of 1000 ppm), area normalisation (against total ion count) and statistical analysis (Wilcoxon/Kruskal–Wallis test). All the spectra were normalised to the total ion current of an m/z value ranging from 2-12,000 and the baselines were subtracted. The part of the spectrum with m/z values less than 2,000 was not used

for analysis because signals generally interfered with peak detection in this area. Peaks with  $m/z$  values between 2,000 and 12,000 were auto-detected with a signal to noise ratio of five.

Hierarchical cluster analysis was performed using the Euclidian and Ward method at 95% explained variance. To evaluate the lowest distance at which a strain would cluster, MALDI-TOF-MS spectra of a single strain from independent repeat runs were included and a similarity distance cut off for strain variability established. From this cut-off point, the different strain groups determined and spectra were imported into Statistica 10 (Stat soft, USA) for model generation and rapid classification. To evaluate the performance of the SVM classification method, data were split into training set 75% and testing set (25%). PCA, (ClinProtTool 3.0; Bruker Daltonics) was used to analyse the proteomic features of the MALDI spectra data from the determined strain groups.

## **5.3 Results**

### **5.3.1 Optimisation of experimental parameters**

Whole-cell MALDI-TOF-MS can readily identify bacteria at species level by detecting a limited number of specific biomarker peaks. Typically, 5 to 10 peaks in the molecular mass range from 2,000 to 11,000 Da have been reported to sufficiently discriminate bacteria at the species level (Dieckmann et al., 2008). At strain level, the requirements related to information content, reproducibility, mass accuracy, and quality of spectra are significantly greater than the requirements for routine whole-cell MALDI-TOF-MS for species identification (Dieckmann et al., 2008). In order to standardise analytical protocol that would minimise these challenges, we evaluated whole-cell MALDI-TOF-MS with intact bacteria and after protein extraction and the influence of the number of replicates on reproducibility. We further evaluated various spectra processing parameters in ClinProTools that would reduce the effect of variability in single strain spectra. These parameters included spectra smoothing, mass range, peak calculation, data reduction and recalibration. Our aim was to determine the simplest procedure that has the potential for automation and that results in MS data with an optimum number of significant picks for strain differentiation.

### 5.3.2 Reproducibility of the MALDI-TOF-MS approach

A high degree of reproducibility is required for any microbiological typing technique to be accepted as valid, especially when there is a need to create a database as is the case with bacteria source tracking. Much of the spectra variation in MALDI-TOF-MS identification of bacteria has been reported to mainly result from the method of sample preparation (Bruyne et al., 2011). We evaluated the reproducibility of MALDI-TOF-MS profiles of samples run from intact bacteria and extracted total cellular proteins (Figure 5.1). Although the intact bacteria cells repeats were comprised of runs carried out over a number of weeks apart, their clustering at SD 18.5 was comparatively lower and consistent than that of MALDI-TOF-MS profiles after protein extraction, whose clustering distance ranged from SD10.1 to 33.4. Furthermore, directly subjecting bacterial to MALDI-TOF-MS analysis resulted in rich peak contents of the spectra and the highest reproducibility when compared to samples run after total protein extraction as previously reported (Hsieh et al., 2008). Therefore, total protein extraction approach was not implemented further. Intact whole-cell MALDI-TOF-MS was used to generate MS spectra for analysis in subsequent studies. Isolates were typed in duplicates to continually validate reproducibility of the typing system,. An isolate was considered correctly typed and identified as a strain if the duplicate profiles belonged to a single strain cluster group.

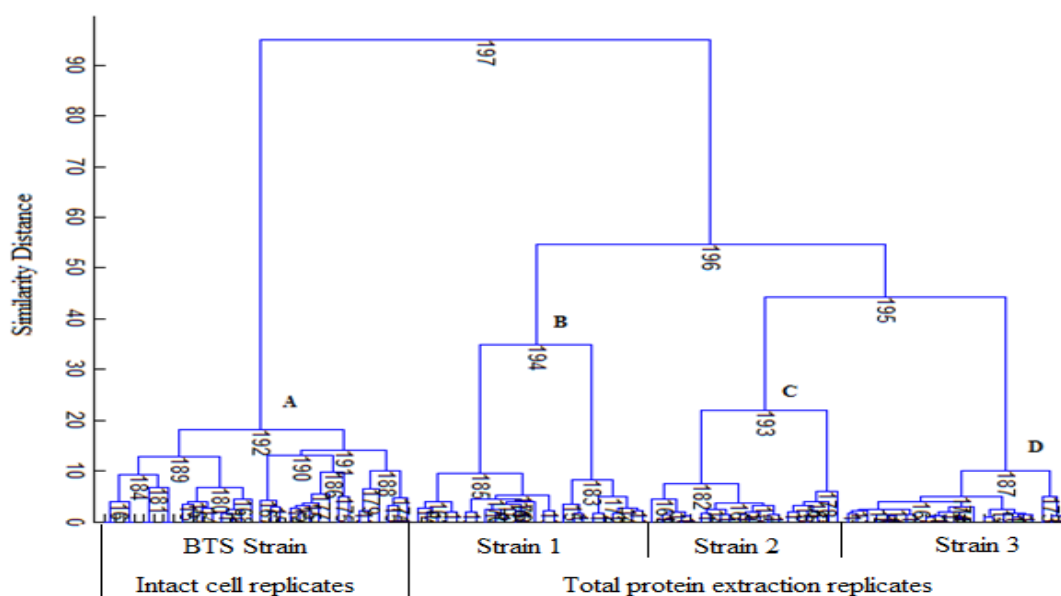


Figure 5. 1: Matrix-assisted laser desorption ionisation time of flight mass spectrometry spectra reproducibility showing (a) Ward Euclidian cluster dendrogram for 4 strain groups spectra obtained with intact bacteria (A) and total protein extraction (B, C and D).

### 5.3.3 Study work flow

During our work with MALDI-TOF-MS spectra of *E. coli* from various sources, we observed various unique spectra profiles among strains from different sources. We therefore determined to create a simple strain identification and characterisation procedure that has the potential for automation and that results in mass spectroscopy MS data with an optimum number of significant peaks for strain differentiation. Initially, spectra of *E. coli* that had been identified with a score of at least 2.000 and PCR confirmed with UidA gene were analysed. These spectra were from river water (67 strains), pigeon faeces (136), sewage (12), roof harvested rainwater (154), humans (53) and bovine milk (5). As a typing system MALDI-TOF-MS has been shown to be equivalent to PFGE, REP PCR and MLST (Novais et al., 2013). Our challenge then was to streamline the strain identification process starting from hundreds of spectra. We employed two approaches: unsupervised hierarchical clustering using PCA data with subsequent classifier selection and visual inspection of the spectra for similarity and direct model construction using the supervised methods including quick classifier, genetic algorithm, supervised neural network, support vector machine and random forest. The performance and reliability of the constructed models were evaluated by splitting the data into training (75%) and test 25% sets and cross-validation across the determined groups.

### 5.3.4 MALDI-TOF-MS hierarchical cluster analysis of *Escherichia coli* isolates

Species specific classification of *E. coli* has been reported by many authors (Novais et al., 2013). The challenge is to create classification systems for strain differentiation which would be applicable in bacterial source training. We evaluated band- and character-based analyses of MALDI-TOF mass spectra for the classification of bacterial strains within *E. coli*. To evaluate the identification potential of this approach, unsupervised hierarchical cluster analysis was applied and the results provided insight into the applicability of the MALDI-TOF-MS data set for strain identification within *E. coli*.

The MALDI-TOF-MS spectra (Bruker Daltonics, Germany) were imported into ClinProTools software for post-processing and generation of proteomic profiles. Normalizing, baseline subtracting, peak defining, recalibrating and comparison of multiple spectra were performed automatically by the ClinProTools software. A total of 506 spectra profiles from 237



environmental *E. coli* isolates and one well characterised strain replicated sixteen times in duplicate were processed. After data pre-processing step, the resulting high quality data set was used for unsupervised clustering using PCA data with 95% explained variance. To evaluate the lowest distance at which a single strain would cluster, 16 replicate spectra of a single stain were included in the dendrogram. The replicate spectra clustered at a SD of 7.34 and the cut-off for unique strains was set at SD 10. At this cut-off distance, a total of thirty-one clusters which represented unique strain groups were selected (Figure 5.2).

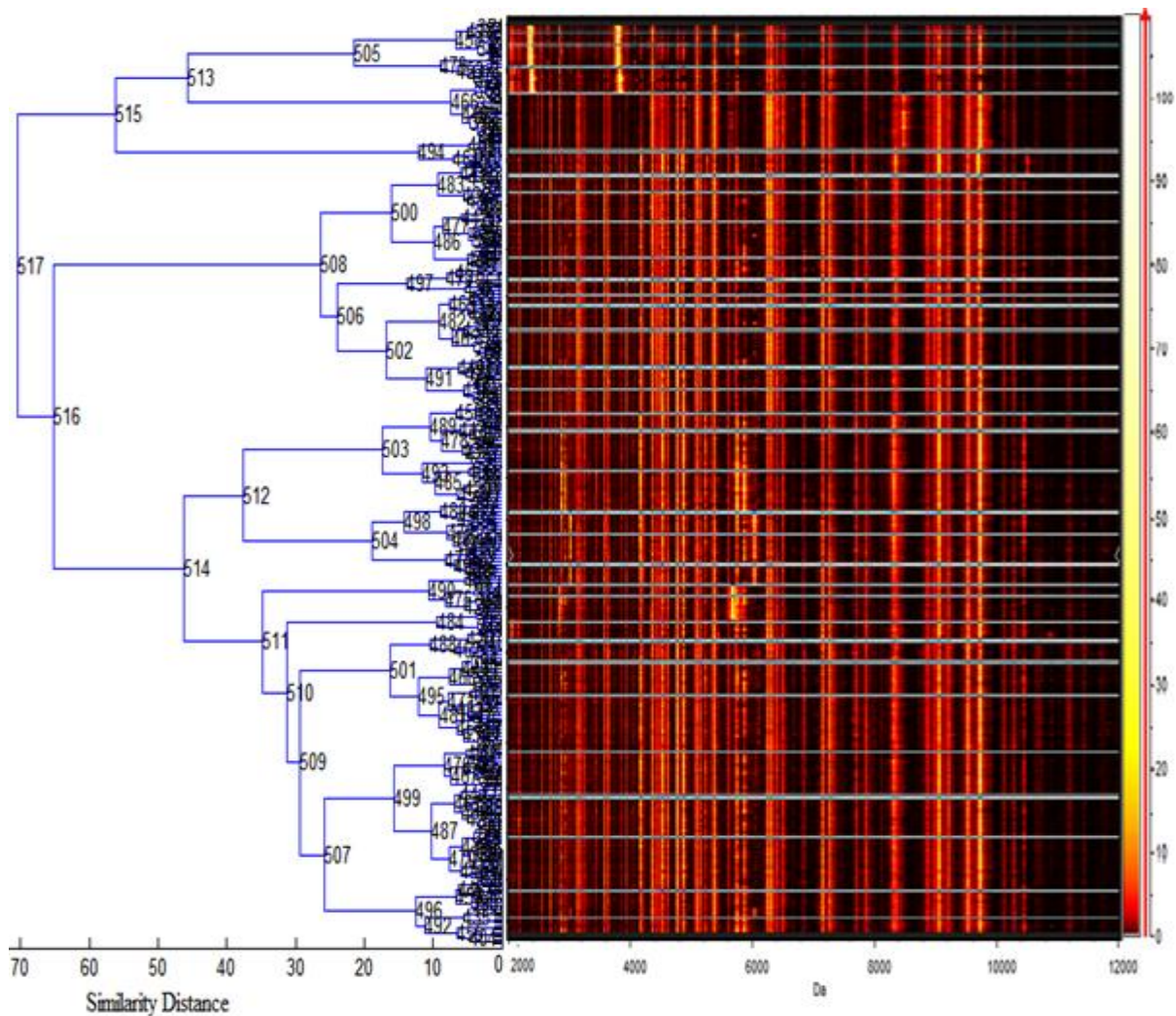


Figure 5. 2: Two-dimensional unsupervised hierarchical cluster analysis of 395 *Escherichia coli* isolates (790 spectra) versus 100 m/z values (rows) dendrogram and associated chromatic gel view of the mass spectra profiles. The normalised expression value for each protein is indicated by a colour with arbitrary units represented by varying shades of brightness. The absolute intensities of the ions are shown on the key to the right and the masses (in kDa) of the ions are shown on the x axis.

The MALDI-TOF-MS spectra of *E. coli* were grouped alongside the respective cluster groups from the generated dendrogram. The dendrogram could be divided into three super-clusters (A, B and C) from which clusters were also divided into sub-clusters. For instance cluster B-2-1 means that, the cluster group belongs to group B under cluster 2 in sub-cluster 1. Four clusters (A-1 to A-4) were observed under group A. These cluster groups had unique peaks that could be observed. Cluster group A-1 and A-2 had highly expressed characteristic peaks between 2300-2400Da and 3700-3880Da. No other cluster groups had similar peak distribution and isolates in these two groups were all from river water. The standard *E. coli* strain replicate profiles included in the analysis clustered in group A-3 was characterised by a highly expressed protein of m/z 6837.29

The fourth cluster group A-4 was characterised by three protein peak mass shifts at 4852.65Da, 7641.19, 9703.28 and 10506.14Da. A total of nine isolates which belonged to this group were all isolated from roof harvested rainwater at a single site. Group B was comprised of four clusters (B-1, B-2, B-3 and B-4) with ten sub-clusters. Unique peak distribution was observed in group B-3-1 and B-3-2, which like group A-4 had peak shifts at position 7641.19Da and 9703.28Da although no peak shift were observed at the 4852.65Da and 10506.14Da position. Group C was comprised of seven clusters with seventeen sub-clusters. Visually apparent peak distributions were observed in cluster groups C-2 and C-3 and their respective sub-clusters. Cluster group C-2 was characterised by peaks 3015.07Da and 6031.48Da in all its sub-clusters. However, sub-cluster group C-2-3 did not have the peak at 2931.30Da and 5866.90Da although it was present in the two other sub-groups (C-2-1 and C-2-2). Cluster group C-3 was characterised by one prominent peak at position 5672.19Da. The rest of the groups did not have visible unique, group distinguishing peaks, although they were recognised to belong to different cluster groups and their respective sub-clusters.

### 5.3.5 Strain groups similarity evaluation

The thirty-one cluster groups that were determined from cluster analysis were considered to be unique strain groups. We therefore determined to evaluate the similarity between these cluster groups. Duplicate profiles of *E. coli* isolate members of the respective cluster groups (strain groups) were imported into MALDI Biotyper 3 (Bruker Daltonics), combined into

mean spectra profiles and clustered together with the reference database isolates (RDI) (Figure 5.3). The MSP cluster dendrogram of observed strains was representative of the dendrogram originally generated by the individual isolates. Closely related clusters were grouped together as closely related strains. Examples of this were strain from clusters B-3-1 and B-3-2; A-1 and A-2; C-1-2 and C-1-3. The generated dendrogram was divided into two primary clusters with our isolates clustering separately from the reference database isolates (RDI). The cluster with reference database isolates had two main sub-clusters with one comprising only of six database reference isolates and the other with two of our strains (A-1 and A-2) and two RDI (*E. coli* w3350 and DHalpha). The cluster with our strains also had two main clusters of which one was comprised of a single RDI (*E. coli* Nissl) whereas the rest of our strains were in the other clusters group in which only 2 RDI (*E. coli* RV412 and B421) were clustered.

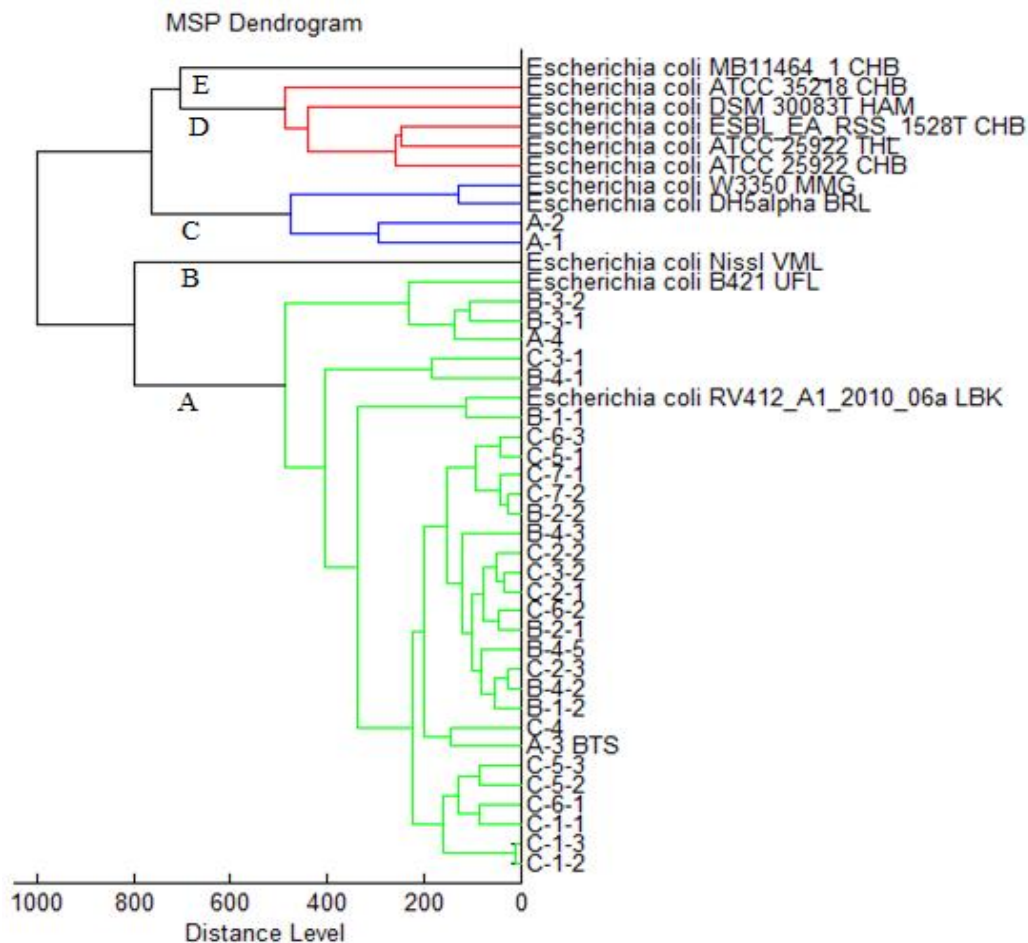


Figure 5. 3: Mean spectrum dendrogram of the thirty-one strain groups identified from two-dimensional unsupervised hierarchical cluster analysis of 395 *Escherichia coli* isolates (790 spectra)

### 5.3.6 Machine learning

Support vector machine data analysis was used to evaluate the uniqueness of the strain groups and to generate classification algorithms that would distinguish between different stain groups.

Table 5. 1: Identification results of a Support Vector Machine model on the binary peak data of the thirty-one unique strain (cluster) groups. The rows denote the observed strain group and the columns the predicted strain group. The actual group membership is shown for in the table. Identification percentages are given at the bottom of the table.

	A-1	A-2	A-3	A-4	B-1-1	B-1-2	B-2-1	B-2-2	B-3-1	B-3-2	B-4-1	B-4-2	B-4-3	B-4-5	C-1-1	C-1-2	C-1-3	C-2-1	C-2-2	C-2-3	C-3-1	C-3-2	C-4	C-5-1	C-5-2	C-5-3	C-6-1	C-6-2	C-6-3	C-7-1	C-7-2
A-	1																														
A-		7																													
A-			1																												
A-				7																											
B-					5																										
B-						8																									
B-							1																								
B-								6																							
B-									4																						
B-										3																					
B-											7																				
B-											1	1																			
B-													5	1																	
B-														7																	
C-															5																
C-																9												1			
C-																	1														
C-																		6													
C-															1				1	6											
C-																					5										
C-																						3	1								
C-																							7								
C-																								5							
C-																									6						
C-																										9					
C-																											1				
C-																												1			
C-																													1		
C-																														8	
C-																															5

Class. Accuracy (%) = 100.000(Train), 89.06(Test), 97.23(Overall) Cross-validation accuracy (%) = 86.77

The SVM classification algorithm determined the rate of cross-validation and recognition capability between the thirty-one strain groups to be 86 % and 100% respectively. The application of SVMs to the binary peak data of cluster groups generated from unsupervised hierarchical cluster analyses gave an indication on how well the technique could discriminate

between the different *E. coli* strains based on MALDI-TOF mass spectra. The results showed a very high strain identification performance. The multi-class confusion matrices, as generated from the identification of the detected strain groups, are reported in Tables 5.1. Each row represents the observed strain group while the columns denote the predicted strain group. The data showed in the table are the actual numbers of isolates that composed each strain group, except for group A-3 which was composed of a single isolate replicate spectrum. The SVM class accuracy was 100 % (Train), 84.36 (Test), 97.23 (Overall) and 86.77% in Cross-validation. Ideally, if all spectra were correctly identified, each strain group should have a 100% identification score. In this case, twenty-three of the strain groups had absolute matches between the observed and predicted classes. Apart from the percentage matches, the table provides information on the misidentified spectra. For example, two single strain group members observed in strain groups C-2-2 were predicted to belonging to strain groups C-1-2 and C-1-3. Misclassifications were observed in eight of the thirty-one observed strain groups. Only two of the strain groups (C-1-2 and C-2-2) had two misclassified isolates in each other whereas the rest of the stain groups (B-3-1, B-4-2, B-4-3, C-1-3, C-2-3 and C-3-1) had single misclassified isolates.

### **5.3.7 Classification of *Escherichia coli* strains by their sources**

After characterising *E. coli* isolates by their profile types and optimizing the classification algorithms, we evaluated the ability of this strain classification system to distinguish the isolates by their sources of origin. The isolates included those collected from RHRW in Pretoria and Johannesburg. Water samples were collected from five sampling areas in Johannesburg (2) and Pretoria (3) and those from urban pigeons, river water, bovine milk, human and the control test strain. Unique grouping by sources was observed for strain group A-4 and B-3-1 with isolates only observed in RHRW site PTA-3. Similar groupings were also observed in clusters B-3-1, B-3-2 and C-2-2 with isolates from RHRW and cluster C-4 with isolates of human origin and clusters 1-1 and A-2 with river isolates only. Isolates from birds had the highest number of shared clusters with those from RHRW in 23 of 30 clusters. Isolates from sewage water classified into two clusters B-1-2 and B-4-1 together with bird and RHRW isolates. Seven of the forty-four human isolates were clustered together with bird and RHRW isolates in strain groups C-1-1, C-1-2 and C-1-5. Similarly isolates from bovine milk were grouped in clusters C-1-3 and C-5-2.

Table 5. 2: Frequency of occurrence of *Escherichia coli* strains from different sources in relation to homology/similarity groups (strains) observed from the dendrogram.

	A-1	A-2	A-3	A-4	B-1-1	B-1-2	B-2-1	B-2-2	B-3-1	B-3-2	B-4-1	B-4-2	B-4-3	B-4-5	C-1-1	C-1-2	C-1-3	C-2-1	C-2-2	C-2-3	C-3-1	C-3-2	C-4	C-5-1	C-5-2	C-5-3	C-6-1	C-6-2	C-6-3	C-7-1	C-7-2	Total
PTA-1	-	-	-	-	1	1	1	-	-	1	-	1	1	2	-	1	-	-	3	-	-	-	-	1	1	2	1	-	-	-	-	17
PTA-2	-	-	-	-	2	-	-	1	-	1	2	3	-	-	-	2	3	-	-	3	-	2	-	-	-	-	-	1	3	2	2	27
PTA-3	-	-	9	-	1	4	2	4	-	1	3	-	-	-	2	2	2	1	-	-	2	-	5	3	10	4	2	2	-	3	62	
JHB-1	-	-	-	1	1	1	-	-	-	-	-	2	1	1	3	2	-	2	-	1	2	-	1	2	3	-	4	2	-	-	29	
JHB-2	-	-	-	1	-	-	1	-	1	1	1	1	2	1	-	-	1	3	-	-	-	-	-	-	-	3	-	1	-	1	2	20
<b>Water total</b>	-	-	-	9	5	3	6	4	4	3	4	8	4	5	2	8	7	3	9	3	1	6	-	7	6	18	5	8	7	3	7	155
Bird	-	-	-	-	4	10	8	5	-	-	9	5	4	8	4	9	5	3	-	4	3	5	-	2	5	9	10	5	9	8	2	136
Sewage	-	-	-	-	6	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12
human	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2	-	-	-	-	-	-	46	3	-	-	-	-	-	-	-	53
Control Strain	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	
Milk	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	1	-	-	-	-	-	5	
River	11	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	
<b>Total</b>	11	7	16	9	9	19	14	9	4	3	19	13	8	13	8	19	16	6	9	7	4	11	46	12	12	27	15	13	16	11	9	395

### 5.3.8 Principal component analysis

Mass spectra of the thirty-one strain groups obtained from unsupervised hierarchical cluster analysis were further subjected to PCA to evaluate how the strain groups were distributed in 3D plane and to establish a mini-optimised profile for *E. coli* strain classification. The state of classification of the thirty-one sets of strain mass spectra was demonstrated in the first three principal components model of PCA (Figure 5.4). The contributions of PC1, PC2, and PC3 to the generation of profile in a percentage plot of the variance explained were approximately 32%, 28%, and 14%, respectively. Figure 5.4 shows the 2 and 3 dimensional loading plots of differentiated peaks generated from PCA analysis. Each point on the loadings graphs represents a peak from the analysed spectra and indicates how principal components are related to the original peaks. The greater the distance of a peak from the origin, the greater its contribution to the variance of the data set. A total of ninety-three differentiated peaks that can serve as potential strain biomarkers were automatically selected by the ClinProTools software. These were then used to determine a minimised differential peak set, that are optimal for *E. coli* strain diversity and their potential application in bacterial source tracking.

Using ClinProTools, t-tests of the peak intensity difference enabled us to rank a series of sixty-five differentiated peaks ( $p < 0.01$ ) from the automatically detected ninety-eight peaks (data not shown). The minimised differential peaks set was determined from the PCA analysis generated factor loadings dimensional view of the differential peak dot. It revealed eleven dots representing

different peaks of mean masses (Da) 2367.87, 2437.24, 2820.54, 3789.31, 3819.61, 3850.16, 5607.24, 5673.52, 5722.87, 6031.68 and 8437.24 (Figure 5.4).

The loading values of each peak are obtained during the calculation of PCs in the loading model (loading 1, loading 2, and loading 3). The peaks are assigned loadings values depending on the amount of variance of a PC they explained, of which the value size is determined by the contribution to the explained variance of a PC. The generated loadings values were between -1 and 1. During the calculation of the PCA algorithm, individual peaks can obtain loading values that originate from the calculation of PCs. In this context, a peak obtains three loading values originating from three PCs (PC1, PC2, and PC3) calculation. Hence, the contribution of single peak through a PC to the mass spectra profile was equal to the product of a principal components contribution and the loading value. The total contribution of an individual peak is therefore a sum of the three contribution values of each peak.

The significance of a peaks contribution to the determined PCA is shown by its distance from the centre in the 2D and 3D view of loading values plots (Figure 5.4). Therefore, loading values of peaks close to the axis are not significant and were consequently excluded from the minimised profile (data not shown). The p values of t-tests and the average value of the eight peaks identified by PCA are shown in Table 5.3. A comparison of the average values of the eleven peaks selected for the minimised profile shows differential expression levels of the selected protein as observed from the peak intensities (Figure 5.5). Peak 2367.87 was only upregulated in strain groups A-1 and A-2 whereas peaks 2437.24, 3819.61 and 3850.16 were only unregulated in group A-2. Similar differential expression were observed for peaks 5673.52 in C-3-2; peak 6031.68 in C-2-1, C-2-2 and C-2-3 and peak 8437.24 in strain group A-3.

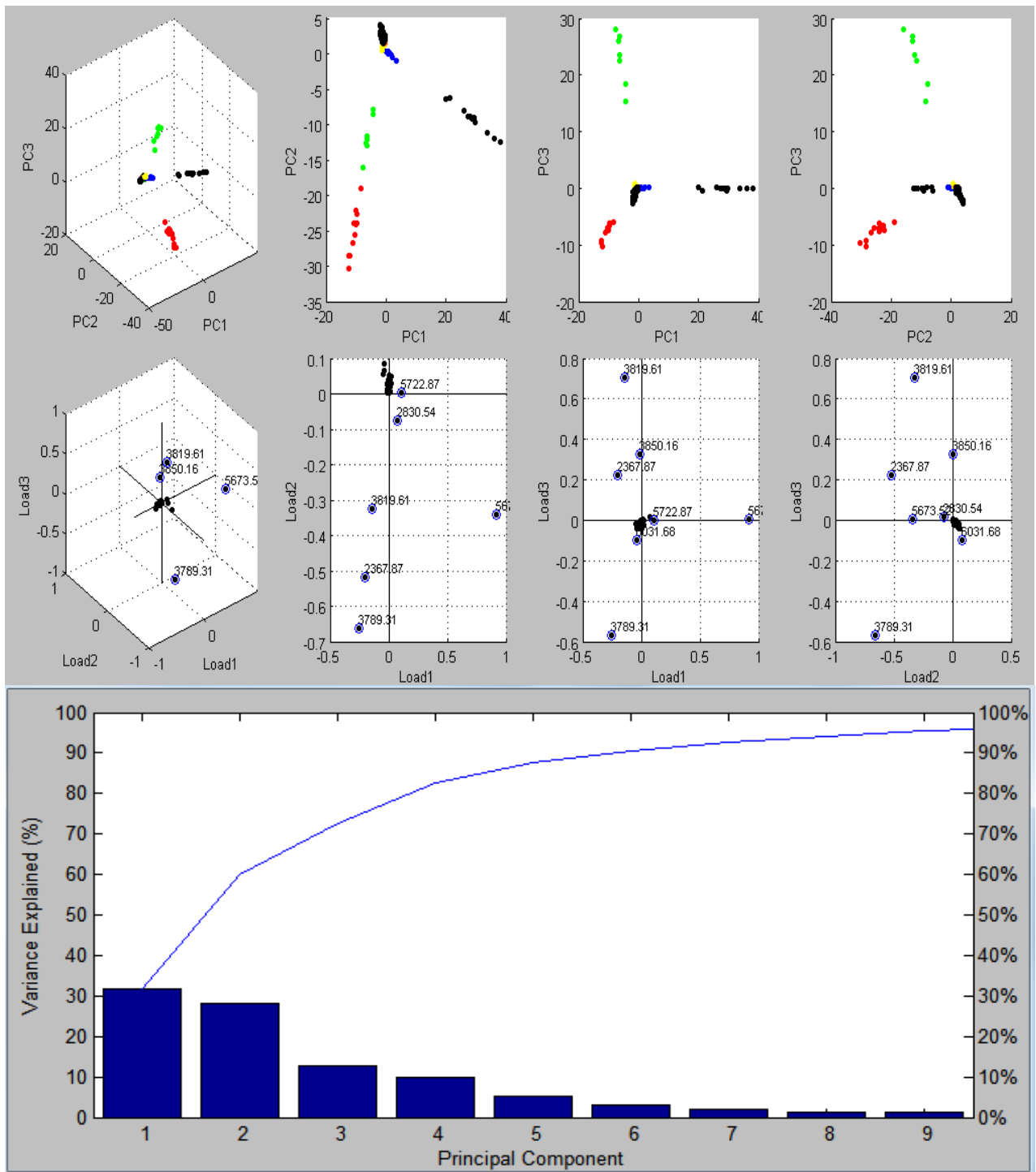


Figure 5. 4: Principal component analysis 3D view and explained variance. The dots show the corresponding spatial scattering state of the differential peaks in loading mode. Each dot indicates the intensity value of a peak; mean masses (Da) 2367.87, 2437.24, 2820.54, 3789.31, 3819.61, 3850.16, 5607.24, 5673.52, 5722.87, 6031.68 and 8437.24 were the differential peaks since its position was away from the centre. The peaks were differential up to the loading value corresponding to the loading (loading1, loading 2, loading 3) model in three binary images.



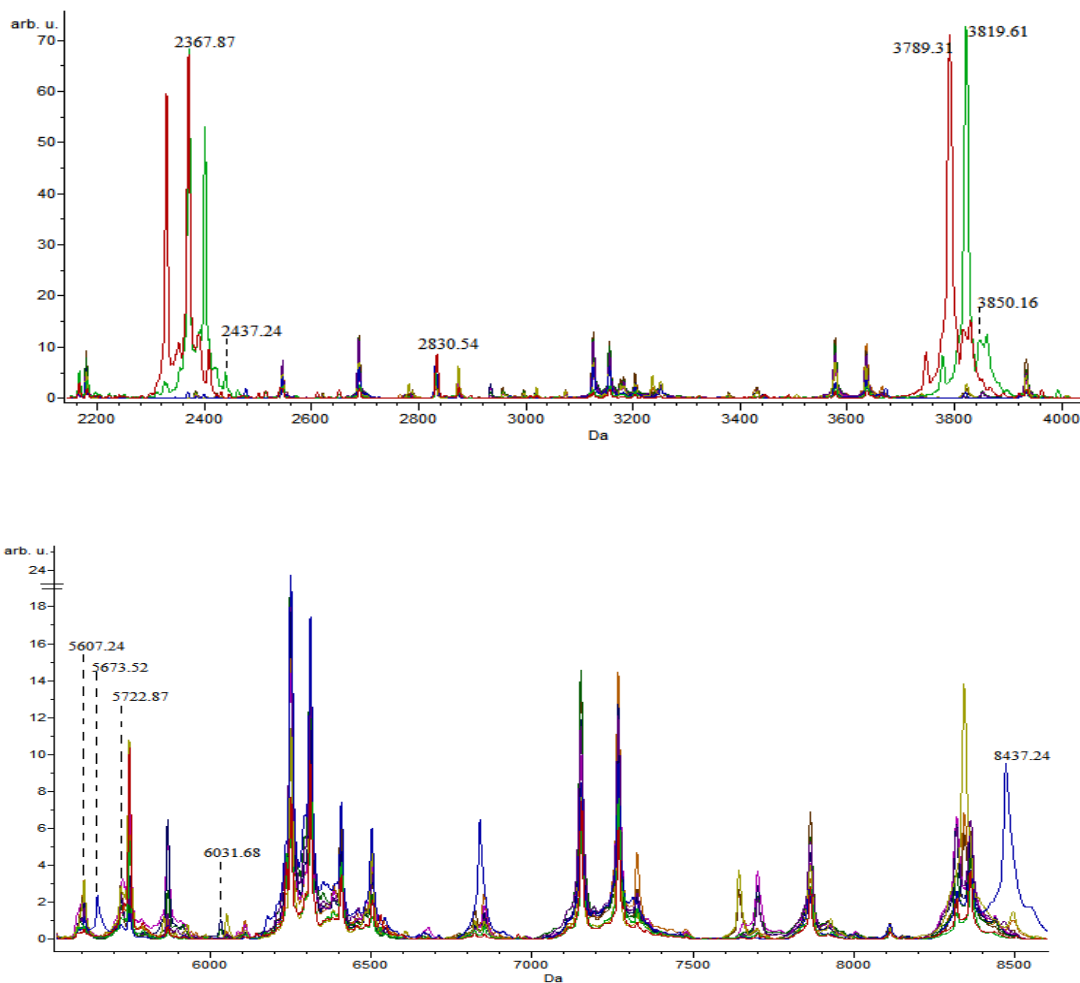


Figure 5. 5: MALDI-TOF mass spectral stack view of selected peaks derived from the original profile according to PCA analysis. The mono-isotopic mass ( $m/z$ ) is shown for each peptide ion peak (A:  $m/z$  2863; B:  $m/z$  1866; C:  $m/z$  4212; D:  $m/z$  2211; E:  $m/z$  2661; F:  $m/z$  5341; G:  $m/z$  4284; H:  $m/z$  5910). Eleven peptides were selected to illustrate group-specific differences in normalised intensities. The differential expression of the peptides can be observed from the peak intensities for the different groups.

To compare the abilities of the minimised profile to classify the thirty-one determined strain groups. The eleven peaks were used to distinguish the strain groups in PCA analysis, unsupervised hierarchical cluster analysis and in support vector machine classification (Figure 5.6). The dendrogram generated from the minimised profiles demonstrated the applicability of minimised profiles. The dendrogram was comparable to the one generated from the sixty-five t-test selected peaks. Closely related strain groups were clustered together and five distinct groups could be observed with the majority of the strain groups falling relatively into closely related one cluster as had been previously determined by PCA analysis. Furthermore, the first three principal components explained about 94% of the observed variance compared to the 74% obtained with

sixty-five t-test determined peaks. An attempt to classify the thirty-one groups using eleven peaks on support vector machine could not generate classification accuracy above 50%. The predicted and observed results were accurate for the four groups (A-1, A-2, A-3 and C-3-2) and consistent misclassifications were observed for the rest of the groups.

Table 5. 3: Data for the intensities of eleven differential peaks obtained from loading plots of the principal component analysis.

Strain Groups	Mass (kDa)										
	5607	2368	3820	2830	5722	3850	3789	5674	6031	8439	2398
A-1	4.68	<b>416</b>	<b>134</b>	<b>48.8</b>	4.8	0.97	<b>694</b>	-4.2	0.27	-34	-20
A-2	9.92	<b>391</b>	<b>565</b>	<b>45.1</b>	6.57	-0.1	<b>-50</b>	-2.4	0.02	-31	<b>218</b>
A-3	10.3	3.38	6.78	<b>36.9</b>	2.73	-0.1	-1.2	<b>29.8</b>	0.42	<b>297</b>	2.82
A-4	32.9	-9.7	22	13.3	<b>12.5</b>	0.42	-0.6	-9.6	-2.7	-84	0.17
B-1-1	21	-9.9	2.56	13	<b>15.5</b>	5.7	-0.9	-6.8	0.28	-98	-1.2
B-1-2	17.9	-13	2.6	<b>28.1</b>	<b>12.2</b>	5.73	0.19	-7.3	4.67	-74	-0.9
B-2-1	24.6	-9.9	3.59	<b>25.4</b>	<b>16.9</b>	6.29	0.37	-7.7	9.1	-88	-0.4
B-2-2	27.3	-5.6	2.44	19.8	<b>14.2</b>	6.7	-0.4	-9.4	0.19	-96	-0.8
B-3-1	36.8	-17	17.7	<b>25.3</b>	<b>21.3</b>	0.67	-2.2	-8.6	0.06	-67	-1.3
B-3-2	38.3	-8.6	15	<b>21.8</b>	<b>15.2</b>	0.68	-0.7	-8.2	0.27	-44	-0.5
B-4-1	29.6	-15	4.61	<b>22.4</b>	<b>22.6</b>	7.2	-1.2	-6.2	16.8	-69	-1.7
B-4-2	28.2	-18	3.58	<b>23.7</b>	<b>18</b>	7.76	-0.7	-7.7	6.07	-63	-1.8
B-4-3	38.8	-20	6.07	19.7	<b>20.7</b>	8.41	-1.2	-10	-1.7	-69	-1.2
B-4-5	28.2	-19	3.34	21.7	<b>16.3</b>	<b>11.3</b>	-0.6	-7.2	11.1	-64	-1.6
C-1-1	10.4	-6	0.01	6.62	9.49	<b>20.8</b>	1.54	-7	-6.5	-90	-0.4
C-1-2	21.1	-7.5	0.59	9.6	8.47	<b>15.6</b>	2.06	-9.7	-4.3	-128	-0.1
C-1-3	20.5	-5.7	1.19	7.79	4.39	<b>12.7</b>	2	-12	-2.8	-89	-0.5
C-2-1	20.4	-5.4	0.64	11	7.39	<b>12.7</b>	1.43	-8.3	<b>248</b>	<b>-100</b>	-1.2
C-2-2	24.4	-8.3	1.25	12.2	<b>10.6</b>	<b>12.8</b>	0.81	-5.6	174	<b>-108</b>	-0.5
C-2-3	30.4	-11	2.78	18.1	<b>14.8</b>	<b>12.2</b>	0.63	-8.9	161	-84	-1.6
C-3-1	20.2	-12	3.69	102	<b>10.6</b>	4.64	-2.1	<b>400</b>	-0.7	-72	-0.4
C-3-2	16.1	-6.3	1.31	85.8	-3.2	6.18	-0.4	<b>374</b>	-0.9	<b>-118</b>	-0.8
C-4	19.6	1.38	-3.4	18.1	3.12	2.56	-2.8	-12	0.02	<b>-100</b>	-0.4
C-5-1	39.2	-11	5.95	17	<b>29.4</b>	7.18	-2.5	-8.4	7.06	-76	-0.5
C-5-2	34.5	-8.6	3.58	13.9	<b>16</b>	<b>10.5</b>	0	-9.4	-0.8	-86	-0.6
C-5-3	36.2	-8.9	4.46	16.6	<b>17.9</b>	8.79	-0.4	-8.5	-1	-87	-0.2
C-6-1	29.2	-12	2.86	13.5	<b>15</b>	<b>14.3</b>	0.29	-7.1	-3.4	-78	-0.4
C-6-2	27.7	-11	2.23	19.4	<b>11</b>	<b>11.4</b>	0.4	-8.7	-0.7	-88	-0.2
C-6-3	28.4	-11	3.04	20.9	<b>15.7</b>	8.45	0.41	-9.2	-1.2	-90	-0.3
C-7-1	19.5	-7.8	1.57	23	<b>11.7</b>	6.34	0.56	-9.4	23.6	-93	-0.9
C-7-2	20.1	-6.8	1.48	14.6	9.65	7.25	0.36	-13	-0.2	-107	-0.8
DAve	34.6	436	569	95.1	32.5	20.9	744	412	255	425	238
PTTA	0	0	0	0	0	0	0	0	0	0	0
PAD	0.317	0	0	0	0.0731	0	0	0	0	0	0

Mass, m/z; PTTA, p-value of t-test (thirty-one classes); PAD, p-value of Anderson-Darling test, which gives information about normal distribution; range 0 to 1; 0: not normal distribution; 1: normal distribution; A-1 to C-7-2: the average values of the respective strain groups.

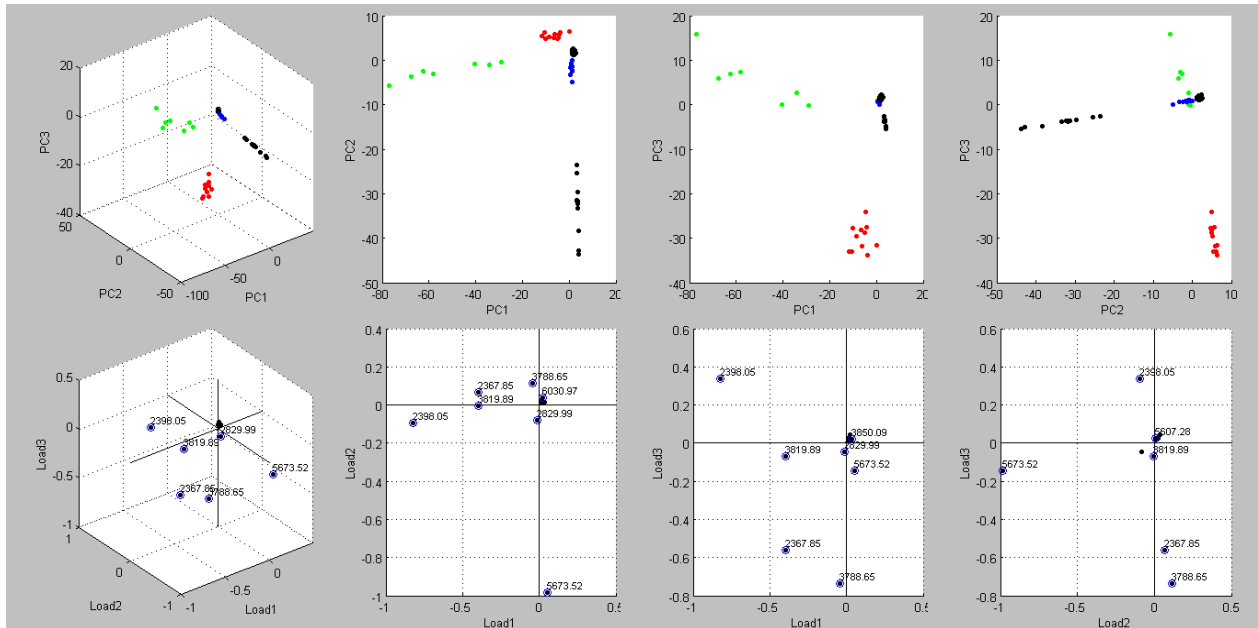


Figure 5. 6: The dimensional image from principal component analysis generated with minimised profile data, showing the distinction between thirty-one *Escherichia coli* strain groups. A: the classification of strain groups in the in the first three principal component model (PC1, PC2, PC3). B: the contribution of eleven principal components to the profiling classification in plot of percentage explained variance of PC. The contributions of PC1, PC2, and PC3 were approximately 45%, 27%, and 22%, respectively.

## 5.4 Discussion

The mitigation of environmental water faecal pollution/contamination requires rapid species identification and strain diversity determination of specific bacteria species which then can be used to identify potential sources of contamination, generally termed bacterial source tracking (Meays et al., 2004). The most commonly used method include phenotypic (biotyping) and genotypic (DNA-based) of which are either time consuming or expensive. Several researches have demonstrated MALDI-TOF-MS to be a time-saving and cost effectiveness method (Novais et al., 2013). Whole-cell MALDI-TOF-MS characterisation and identification of microbial species is generally achieved by comparison of experimental mass data for a set of intact protein ions desorbed from whole bacterial cells with a database of reference spectra (Bruyne et al., 2011). Typically up to thirty constant peaks are detected in mass spectra obtained and are predominantly in the molecular mass range from 4,000 to 13,000 Da (Vargha et al., 2006). This number of peaks has been shown to be sufficient for bacterial identification at the genus and species levels of pathogens, including *E. coli*, *Campylobacter* spp., *Salmonella* spp., *Pseudomonas* spp., *Yersinia* spp., and *Listeria* spp. from both clinical

and environmental sources (Carbonnelle et al., 2007; Dieckmann et al., 2008; Romanus et al., 2011; Cho et al., 2012).

In this study we applied MALDI-TOF-MS to the characterisation and differentiation of *E. coli* strains isolated from RHRW and pigeon faeces, river water, sewage water, bovine milk and humans, in an attempt to evaluate the applicability of MALDI-TOF-MS in distinguishing *E. coli* strains and possible application in BST. For each isolate, MALDI-TOF-MS analysis yielded a spectrum representing a large set of bacterial products. Limited variation between duplicate spectra was observed, indicating that the method was robust and reproducible for our experimental conditions. Peak intensity were low for masses greater than 13 kDa, probably due to low ionisation yields or poor solubility for higher mass proteins (Sauer et al., 2008). Nevertheless, the number, quality, and variation of the peaks detected were sufficient to reproducibly distinguish bacterial species and demonstrate the potential diagnostic value of this approach.

Whole-cell MALDI-TOF-MS generates numerical data (mass peaks). Those data are easy to exchange between laboratories, compared to molecular fingerprinting methods such as PFGE, which is currently accepted as the gold standard method for bacterial typing (Siegrist et al., 2007). Therefore, it can be compared with DNA sequence data handling used, for example, in MLST. However, whole-cell MALDI-TOF-MS assesses the allelic variation in multiple genes (mainly house-keeping genes) in a strain by determining variations on protein level, in this case, mass variations (Murray, 2010). Silent mutations are therefore not assessed and due to the fact that such variations accumulate very slowly in-house-keeping genes, the discriminative power might be generally lower than for MLST (Sauer et al., 2008).

In order to differentiate bacteria at levels below the species level, spectra with a high number of reproducible protein peaks are required. Fingerprint-based approaches for subtyping bacteria at levels below the species level tend to be less useful than approaches used for species identification, primarily because of the high overall similarity of MS fingerprints within species and the difficulty of reproducibility and detecting sufficient numbers of biomarkers with specificities below species level (Albrethsen, 2007). Promising results with

respect to taxonomic resolution were obtained by Siegrist et al. (2007) who demonstrated the ability of MALDI-TOF-MS to effectively differentiate strains and group environmental isolates of *E. coli* according to their respective sources than rep PCR. MALDI-TOF-MS detects a large spectrum of proteins and should therefore, theoretically be able to discriminate between different strains of the same species (Murray, 2010). A large proportion of MALDI-TOF-MS detected proteins and fragments have a ribosomal origin and sceptical concerns have been raised concerning its claimed taxonomic resolution. Yet, despite the highly conserved nature of ribosomal proteins, slight sequence variations can occur even at strain level (Freiwald and Sauer, 2009).

An overall mean similarity between duplicate MALDI-TOF-MS profiles of  $96.85 \pm 2.10\%$  was observed. These results imply good reproducibility and are in agreement with results from a previous study (Mellmann et al., 2009). However, poor reproducibility of MALDI-TOF-MS has also been reported by other authors (Siegrist et al., 2007). Reasons for this could be the methods of cell preparation, the different mass ranges used and the matrix. Since MALDI-TOF-MS is a chemotaxonomic technique, several parameters, including medium and cell age can significantly affect its reproducibility (Ruelle et al., 2004). Generally, species level identification is not significantly hampered by these experimental differences although resolution at subspecies and strain level might become difficult (Freiwald and Sauer, 2009). However, culture medium, time period of cultivation and preparation protocols have been reported not to interfere in species, subspecies and strain differentiation (Barbuddhe et al., 2008; Seibold et al., 2010). Consequently if the technique is standardised, and measured in a mass range of 2 to 20 kDa, high inter-laboratory reproducibility and robustness under different culture conditions can be achieved (Barbuddhe et al., 2008; Mellmann et al., 2009).

Since organisms are best analysed from the media on which they grow best, cultivation on alternative media might initiate the production of stress-induced compounds and generate extra peaks in the MALDI-TOF-MS profile (Barbuddhe et al., 2008). This, however, is not significant since stress response systems show a high degree of similarity in prokaryotes. Furthermore, at species to strain taxonomic levels bacteria react in a similar way and hence

generate spectral profiles that will not be differentiated based on stress response (Vargha et al., 2006).

#### **5.4.1 Principal component analysis**

In general, bacteria are highly complex systems that consist of lipids, peptides, proteins, and salts. The MALDI mass spectra obtained from the analysis of whole cells provide information regarding these types of molecules (Bruyne et al., 2011). MALDI mass spectra are a type of multivariate data, with each mass signal defining one molecular dimension. Hence, multivariate statistical methods can be used to evaluate differences in mass spectra. PCA is one of the most widely used multivariate statistical methods designed to extract variance within a data set (Shao et al., 2011). PCA reduces the dimensionality of the data set while retaining the information present in the original data set. The dimensionality of the data set can be reduced to 2D or 3D coordinate system and the spectrum is represented by a point, spectra with similar variation characteristics can then be clustered together and the differences between sample groups can be readily visualised in the system (Cho et al., 2012).

In ClinProtTools software, we used the genetic algorithm (GA) and ANN analysis schemes which selected a subset of relevant peaks to establish discrimination models peak intensities. These are, however slow and were further complicated by our large dataset. The use of more complicated statistical algorithms including SVM brought improvement of algorithms resulting in a high recognition rate, although it ignored usefully differentiated peaks that may represent a potential biomarker (Cho et al., 2012). The whole profile, comprising of too many subtly differentiated peaks, may be inconvenient for rapid biomarker detection critical for the particular classification under consideration (Shao et al., 2011). It is therefore imperative that a mini-optimised profile, comprising fewer peaks than the original using a simple statistical algorithm, be established. This is quite relevant as discriminating picks may be associated with specific biochemical functions, which will aid in the characterisation and understanding of the observed diversity. For this purpose, we therefore used PCA in ClinProtTools software to generate a mini-profile and choose the potential biomarker peaks that would explain the observed variation in the thirty-one cluster groups.

The thirty-one cluster groups were loaded into the software and the result was demonstrated on the PCA image. The dimensional view of the first three loading values model from PCA (Loading1, Loading2, Loading3) is shown in Figure 5.4. Eleven black dots standing for the different m/z peaks were selected as the mini-profile peaks according to loading values and SC values (Table 5.3). The most significant peaks were identified by their distance from the centre (Figure 5.4). Although a list of the most significant peaks was obtained using the t-test, we did not consider them but rather the eleven discriminating peaks selected by the PCA algorithm. Previous reports on the use of t-test and PCA in biomarker discovery have revealed that PCA derived peaks gave a better classification than those derived from the p values in t-test (Shao et al., 2011). These findings point to the effect that differential peaks obtained from PCA possessed much better classification ability than those from t-test. This can be explained by the fact that the PCA algorithm referred to more factors such as down-regulation, than t-test. PCA analysis generated fewer peaks (n=11) than those generated by p values (n=53) from t-test. Hence PCA data is easy to analyse in view of the small number of the peaks and is a better method than that based on p values, where too many peaks are significantly different ( $p < 0.01$ ).

The focus of our study was not on biomarker discovery but on the ability of PCA to create a mini-optimised profile to discriminate *E. coli* strains. We therefore did not focus on the loading contribution of the various PCA detected peaks. However the observed peaks had unique distributions in which case some were upregulated and/or down-regulated in some strain groups. PCA analysis was able to discriminate the thirty-one strain groups into five lineage groups. Four of these lineages were comprised of single strain groups whereas the fifth lineage was comprised of the remaining twenty-seven strain groups. The member clusters in the common lineage included isolates from humans, sewerage water, bovine milk, birds and RHRW. Strains in the clusters A-1 to A-4 could be considered rare while the prevalence of a common lineage with minor variants is consistent with the observation that *E. coli* are generally clonal in nature and exhibit limited variation in most environments. While strain groups in the common lineage could not be distinguished in 3D PCA view, the dendrogram constructed using 95% explained variance data reveal the grouping nature within the common lineage

### 5.4.2 Support vector machine

The SVM data analyses were performed with the aim of evaluating the discriminatory effect in the data set. This kind of analyses gives an indication on how well the technique can discriminate between the MALDI-TOF mass spectra of the different strains. The application of SVMs to the binary peak data and the test results showed a very high strain identification performance. The misclassification observed in the strain groups suggest that these could be closely related strain groups which could potentially be considered as one. In developing a strain identification scheme, it is important to evaluate the classification scheme based on the calculation of a global performance measure. The calculated accuracy values here corresponds to the training, testing and cross-validation which sum all the correct predictions (main diagonal of matrix) to the data set size. However our calculated accuracy metric value is somewhat biased due to the imbalanced nature of the sample sizes. Although not calculated here, average F-score values calculated for each strain group comparison would give a better estimation on the identification performance of the technique (Bruyne et al., 2011). Findings show that MALDI-TOF-MS analysis, combined with thorough pre-processing and intelligent prediction models can raise the identification performance of *E. coli* strains.

The MALDI-TOF-MS analysis has been shown here to have a very high identification capacity among *E. coli* species and SVM take advantage of the peak patterns embedded in the MALDI-TOF mass spectra. This SVM approach therefore enables the extension of MALDI-TOF-MS spectra in bacterial characterisation.

The SVM enables the identification of strains in a range of seconds, are easy to handle, and are fast with no time limiting constraints during retraining and construction of models (Bruyne et al., 2011). A number of models that could equally perform these classifications including GA and Artificial Neural Network (ANN) could be implemented. However these are slow and could take days to run. Speed of execution is a critical aspect that is important in order to cope with the ever-increasing changing microbial taxonomy and a growing amount of data, emanating from the high-throughput capability of the MALDI-TOF-MS analysis technique.



Storing MALDI-TOF-MS data from diverse strains of a single species origination from diverse environments would be an ideal starting point for the development of bacterial source tracking system. However, the currently existing database focuses on species identification than strains and includes the commercially available BioTyper and Saramis databases. The current exploited identification systems by these databases is achieved by mass spectra entry of an unknown isolate and matching against the corresponding database with the result given as a matching score (Bruyne et al., 2011). The SVM approach used in our analysis differ from the one applied in these commercial databases. Our analysis approach exploits all spectral information and variability in the data sets to construct an identification model, and consequently, allow the identification unknown MALDI-TOF-MS profiles. Building from the already known high species resolution of the MALDI-TOF-MS analysis, we assume the same variability exists in the strains within the identified species. Given, the rapid and extensive data generation and the differentiation power the intelligent machine learning techniques, it is apparent that that the combining mass spectra analysis and SVM learning aspect results in highly efficient and effective tool for identification of bacterial strain with a species (Sariyar-akbulut, 2009; Shao et al., 2011).

### **5.4.3 Comparison of MALDI-TOF-MS analysis and REP PCR analysis**

The taxonomic resolution of both REP PCR and MALDI-TOF-MS allows for strain level identification. Although REP PCR has conventionally been used to this end, the application of MALDI-TOF-MS to characterise bacterial has already been described (Novais et al., 2013). However, REP PCR cannot be applied with confidence on species level identification. Furthermore, it is difficult to create a universal standard hence inter-laboratory database comparisons are difficult to implement. MALDI-TOF-MS circumvent most of the challenges with REP PCR. The handling of large number of bacteria as encountered in bacterial source tracking, may require that the technique be automated. Whereas REP PCR requires genomic DNA, and therefore an extra cultivation step, DNA extraction, PCR and gel electrophoresis (Vechia et al., 2003). MALDI-TOF-MS can directly be applied on a single colony. Furthermore, comparing the workload and speed, MALDI-TOF-MS is evidently the most suited technique for fast automated strain identification and differentiation. Successful inter-laboratory analyses have already been reported for MALDI-TOF-MS analysis. The availability of commercial database systems for species identification that rely on the inter-

laboratory comparison of spectra is a further advantage to MALDI-TOF-MS based bacterial characterisation.

Taken a step further, one legitimate question would be why should we not then employ MLST for which we know that the reproducibility of sequences is better than that for mass spectra, since ion intensities can fluctuate among spectra (Shao et al., 2011; Cho et al., 2012). Although MLST might offer a better system, it involves the further cultivation and DNA extraction like in REP PCR. Furthermore, prior sequence knowledge is required to design and optimise primer sequences whereas MALDI-TOF-MS on the contrary does not require prior knowledge before the analysis. Due to the potential of ion intensity fluctuate among spectra, one should therefore not consider ion intensities but rather their intensity ratio should be taken into account. This approach will limit the challenges of reproducibility in MALDI-TOF-MS systems (Fox et al., 2010).

## 5.5 Conclusion

MALDI-TOF-MS is a valuable tool for the rapid screening and identification of bacterial strains. The minimal sample preparation involved with whole-cell MALDI-TOF-MS and the speed of the data acquisition together with its potential for high-throughput sample processing automation, makes it a better alternative to the currently employed relatively more labour intensive and expensive techniques. The MALDI-TOF-MS spectra produced from whole-cell have taxonomically relevant characteristic features that can be implemented in differentiation of bacteria at genus, species and subspecies levels, although only a small proportion of the bacterial proteome is detected. In the present study, whole-cell MALDI-TOF-MS was used for the identification and strain differentiation of *E. coli* isolates. Unsupervised hierarchical analysis generated thirty-one strain groups. Principal component analysis of these strain groups illustrated the presence of strain specific information. Data were first analysed based on noise score calculation from which spectra were either accepted or rejected for further analysis. The processed spectra together with the generated data sets were successfully used for strain identification within the *E. coli* species. Numerical analysis of (nearly all) spectra, obtained resulted in strain specific clustering that could be implemented in support vector machine. The implementation of the machine learning

technique (support vector machine) for the recognition and identification of the various *E. coli* strains from MALDI-TOF mass spectra proved to be very successful. The use of whole-cell MALDI-TOF-MS spectra coupled to pre-processing techniques and automated machine learning techniques provides a rapid and relatively simple method for the identification and characterisation of any bacterial species.

## 5.6 References

- Ahmed, W., T. Gardner, and S. Toze. 2011a. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Ahmed, W., L. Hodgson, N. Masters, J. P. S. Sidhu, M. Katouli, and S. Toze. 2011b. Occurrence of intestinal and extraintestinal virulence genes in *Escherichia coli* isolates from rainwater tanks in Southeast Queensland, Australia 77: 7394–7400.
- Albrethsen, J. 2007. Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clinical Chemistry* 53: 852–858.
- Barbuddhe, S. B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty, and T. Hain. 2008. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionisation time of flight mass spectrometry. *Applied and Environmental Microbiology* 74: 5402–5407.
- Benagli, C., V. Rossi, M. Dolina, M. Tonolla, and O. Petrini. 2011. Matrix-assisted laser desorption ionisation time of flight mass spectrometry for the identification of clinically relevant bacteria. *PLoS One* 6: e16424.
- Bittar, F., Z. Ouchenane, F. Smati, D. Raoult, and J. -M. Rolain. 2009. MALDI-TOF-MS for rapid detection of staphylococcal Pantone-Valentine leukocidin. *International Journal of Antimicrobial Agents* 34: 467–470.
- Bruyne, K. De, B. Slabbinck, W. Waegeman, P. Vauterin, B. De Baets, and P. Vandamme. 2011. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and Applied Microbiology* 34: 20–29.
- Carbonnelle, E., C. Mesquita, E. Bille, N. Day, B. Dauphin, J. Beretti, A. Ferroni, L. Gutmann, and X. Nassif. 2011. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clinical Biochemistry* 44: 104–109.

- Carbonnelle, E., J. -L. Beretti, S. Cottyn, G. Quesne, P. Berche, X. Nassif, and A. Ferroni. 2007. Rapid identification of staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionisation time of flight mass spectrometry. *Journal of Clinical Microbiology* 45: 2156–2161.
- Carlos, C., F. Alexandrino, N. C. Stoppe, M. Inês, Z. Sato, and L. M. M. Ottoboni. 2012. Use of *Escherichia coli* BOX-PCR fingerprints to identify sources of fecal contamination of water bodies in the State of São Paulo, Brazil. *Journal of Environmental Management* 93: 38–43.
- Cho, Y., Y. Chiang, J. Shiea, and M. Hou. 2012. Combining MALDI-TOF and molecular imaging with principal component analysis for biomarker discovery and clinical diagnosis of cancer. *Genomic Medicine, Biomarkers and Health Sciences* 4: 3–6.
- Decristophoris, P., A. Fasola, C. Benagli, M. Tonolla, and O. Petrini. 2011. Identification of *Staphylococcus intermedius* Group by MALDI-TOF-MS. *Systematic and Applied Microbiology* 34: 45–51.
- Dieckmann, R., E. Strauch, and T. Alter. 2010. Rapid identification and characterisation of *Vibrio* species using whole-cell MALDI-TOF mass spectrometry. *Journal of Applied Microbiology* 109: 199–211.
- Dieckmann, R., R. Helmuth, M. Erhard and B. Malorny 2008. Rapid classification/identification of salmonellae at the species and subspecies level using whole-cell MALDI-TOF mass spectrometry. *Applied and Environmental Microbiology* 74: 7767–7778.
- Field, K., E. Chern, L. Dick, J. Fuhrman, J. Griffith, P. Holden, M. LaMontagne, B. Olson, and M. Simonich. 2003. A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. *Journal of Water and Health* 1: 181–194.
- Fox, K., A. Fox, J. Rose, and M. Walla. 2011. Speciation of coagulase negative staphylococci, isolated from indoor air, using SDS PAGE gel bands of expressed proteins followed by MALDI-TOF-MS and MALDI-TOF-TOF MS-MS analysis of tryptic peptides. *Journal of Microbiological Methods* 84: 243–250.
- Fox, K., A. Fox, T. Elbner, C. Feigley, D. Salzberg, and T. Elssner. 2010. MALDI-TOF mass spectrometry speciation of staphylococci and their discrimination from micrococci

- isolated from indoor air of schoolrooms. *Journal of Environmental Monitoring* 12: 917–923.
- Freiwald, A., and S. Sauer. 2009. Phylogenetic classification and identification of bacteria by mass spectrometry. *Nature Protocols* 4: 732–742.
- Giebel, R. A., W. Fredenberg, T. R. S. Ã, and T. R. Sandrin. 2008. Characterisation of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption / ionisation time-of-flight mass spectrometry. *Water Research* 42: 931–940.
- Harwood, V., B. Wiggins, C. Hagedorn, R. Ellender, J. Gooch, J. Kern, M. Samadpour, A. Chapman, B. Robinson, and B. Thompson. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *Journal of Water and Health* 1: 153–166.
- Harz, M., P. Rösch, K. -D. Peschke, O. Ronneberger, H. Burkhardt, and J. Popp. 2005. Micro-Raman spectroscopic identification of bacterial cells of the genus *Staphylococcus* and dependence on their cultivation conditions. *The Analyst* 130: 1543–1550.
- Ma, H. J., L. L., Fu, and J. R. Li. 2011. Differentiation of fecal *Escherichia coli* from human, livestock, and poultry sources by rep PCR DNA fingerprinting on the shellfish culture area of East China Sea. *Current Microbiology* 62: 1423–1430.
- Hsieh, S., C. Tseng, Y. Lee, A. Kuo, C. Sun, Y. Lin, and J. Chen. 2008. Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS. *Molecular and Cellular Proteomics* 7: 448–456.
- Lay, J. O. 2001. MALDI-TOF Mass spectrometry of bacteria. *Mass Spectrometry Reviews* 20: 172–194.
- Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder. 2004. Source tracking fecal bacteria in water: A critical review of current methods. *Journal of Environmental Management* 73: 71–79.
- Mellmann, A., F. Bimet, C. Bizet, A. D. Borovskaya, R. R. Drake, U. Eigner, A. M. Fahr, Y. He, E. N. Ilina, M. Kostrzewa and D, Harmsen. 2009. High inter-laboratory reproducibility of matrix-assisted laser desorption ionisation time of flight mass

- spectrometry-based species identification of nonfermenting bacteria. *Journal of Clinical Microbiology* 47: 3732–3734.
- Murray, P. R. 2010. Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry: usefulness for taxonomy and epidemiology. *Clinical Microbiology and Infection* 16: 1626–1630.
- Myoda, S., C. Carson, J. Fuhrmann, B. Hahm, P. Hartel, H. Yampara-Iquise, L. Johnson, R. Kuntz, C. Nakatsu, M. Sadowsky, and M Samadpour. 2003. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. *Journal of Water and Health* 1: 167–180.
- Neuhof, T., R. Dieckmann, I. S. Druzhinina, C. P. Kubicek, and H. Von Dohren. 2007. Intact-cell MALDI-TOF mass spectrometry analysis of peptaibol formation by the genus *Trichoderma/Hypocrea*: can molecular phylogeny of species predict peptaibol structures? *Microbiology* 153: 3417–3437.
- Novais, Â., C. Sousa, Fernandez-Olmos, A. Lopes, Gonzalez, J. Coque, T. M. Moreno, R. Canton, and L. Peixe. 2013. Discrimination of major *Escherichia coli* clones by matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS). In 23rd European Congress of Clinical Microbiology and Infectious Diseases (ECCMID). Berlin, Germany.
- Prod'hom, G., A. Bizzini, C. Durussel, J. Bille, and G. Greub. 2010. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for direct bacterial identification from positive blood culture pellets. *Journal of Clinical Microbiology* 48: 1481-1483.
- Romanus, I. I., A. T. Eze, O. A. Egwu, A. N. Felicitas, and N. A. Chidiebube. 2011. Comparison of matrix-assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF MS) with conventional culture and biochemical method of bacteria identification to species level. *Journal of Medical Laboratory and Diagnosis* 2: 1–4.
- Ruelle, V., B. El Moulaj, W. Zorzi, P. Ledent, and E. De Pauw. 2004. Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 18: 2013–2019.

- Sariyar-akbulut, B. 2009. Rapid differentiation of new isolates with MALDI-TOF mass spectrometry via discriminant function analysis based on principal components. *Korean Journal of Chemical Engineering* 26: 1645–1651.
- Sauer, S., A. Freiwald, T. Maier, M. Kube, R. Reinhardt, M. Kostrzewa, and K. Geider. 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One* 3: e2843.
- Seibold, E., T. Maier, M. Kostrzewa, E. Zeman, and W. Splettstoesser. 2010. Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionisation-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels identification of francise. *Journal of Clinical Microbiology* 48: 1061–1069.
- Shao, C., Y. Tian, Z. Dong, J. Gao, Y. Gao, and X. Jia. 2011. The use of principal component analysis in MALDI-TOF MS: A powerful tool for establishing a mini-optimised proteomic profile. *American Journal of Biomedical Sciences* 4: 85–101.
- Siegrist, T. J., P. D. Anderson, W. H. Huen, G. T. Kleinheinz, C. M. McDermott, and T. R. Sandrin. 2007. Discrimination and characterisation of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS). *Journal of Microbiological Methods* 68: 554–562.
- Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner. 2002. Microbial source tracking: state of the science. *Environmental Science and Technology* 36: 5279–5288.
- Stoeckel, D. M., M. V Mathes, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. L. O'brien, T. W. Fenger, M. Samadpour, K. M. Strickler, T. L. O. Brien, and B. A. Wiggins. 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. *Environmental Science and Technology* 38: 6109–6117.
- Tong, D. L., D. J. Boocock, C. Coveney, J. Saif, S. G. Gomez, S. Querol, R. Rees, and G. R. Ball. 2011. A simpler method of pre-processing MALDI-TOF MS data for differential biomarker analysis: stem cell and melanoma cancer studies. *Clinical Proteomics* 8: 14.
- Vargha, M., Z. Takáts, A. Konopka, and C. H. Nakatsu. 2006. Optimisation of MALDI-TOF MS for strain level differentiation of *Arthrobacter* isolates. *Journal of Microbiological Methods* 66: 399–409.

Vechia, D., G. Corc, and L. Gustavo. 2003. Characterisation and genetic diversity via REP PCR of *Escherichia coli* isolates from polluted waters in southern Brazil. *FEMS Microbiology Ecology* 45: 173–180.



## CHAPTER SIX

### ANTIBIOTIC RESISTANCE IN *ESCHERICHIA COLI* ISOLATES FROM ROOF HARVESTED RAINWATER (RHRW) TANKS

#### Abstract

The objective of this study was to investigate the risks associated with the use of roof harvested rainwater RHRW and the implication of pigeons as the most likely source of contamination by testing for antibiotic resistance profiles of *E. coli*. A total of 239 *Escherichia coli* were isolated from thirty fresh pigeon faecal samples (130 isolates), 11 RHRW tanks from three sites in Pretoria (78) and two in Johannesburg (31). In all samples, resistance to ampicillin (27.9%), gentamicin (23.6%), amikacin (24%), tetracycline (17.4) and amoxicillin (16.9%) were the most frequently encountered form of resistance. However, a relatively higher proportion of isolates from pigeon faeces were antibiotic resistant than those from RHRW. The highest number of phenotypes was observed for single antibiotics and no single antibiotic resistance was observed for chloramphenicol, ceftriaxone, gentamicin, ceftiofur, cotrimoxazole, although they were detected in multiple antibiotic resistance (MAR) phenotypes. The highest multiple antibiotic resistance (MAR) phenotypes were observed for a combination of four antibiotics, on isolates from JHB (18.8%), pigeon faeces (15.2%) and Pretoria (5.1%). The most abundant resistance phenotype to four antibiotics, Ak-Gm-Cip-T was dominated by isolates from pigeon faeces (6.8%) with Pretoria and Johannesburg isolates having low proportions of 1.3% and 3.1%, respectively. Future studies should target isolates from various environmental settings in which rainwater harvesting is practiced and the characterisation of the antibiotic resistance determinant genes among the isolates.

Keywords: Antibiotic resistance, *E. coli*, rainwater, contamination, pigeon faeces

## 6.1 Introduction

Rainwater harvesting (RWH) at the household level is increasingly advocated for in many countries as an environmentally friendly and sustainable alternative water source (United Nations Environment Programme and Stockholm Environment Institute, 2009). In under-developed and developing countries rainwater harvesting is particularly promising in reducing the number of people lacking adequate water supplies given that reticulated water is either financially or technically impractical. Although rainwater is generally considered clean, the potential of public health risk associated with the presence of pathogenic microorganisms cannot be ignored (Ahmed et al., 2008; 2010). Faeces of wild birds, insects, mammals, and reptiles that have access to the roof can be washed into the holding tank during rain events. Consequently, contamination of harvested rainwater by enteric bacteria including faecal coliforms, *Escherichia coli*, and *Enterococcus* spp. (Spinks et al., 2006; Sazakli et al., 2007) which are commonly found in the guts of warm blooded animals has been reported (Sung et al., 2010; Lee et al., 2010; Ahmed et al., 2011a, 2011b). These bacteria have a wide natural host range, which includes all warm blooded animals, some cold-blooded animals, and environmental reservoirs, such as sediments and free-living strains (Harwood et al., 1999; Whitman and Nevers, 2003; Power et al., 2005).

Although several enteric bacteria can be used to monitor faecal contamination, *E. coli* is the most used indicator bacteria as well as a regulatory organism of faecal pollution in aquatic environments. *Escherichia coli* density at elevated levels, primarily provide evidence of faecal pollution and secondarily reflect the possible presence of bacterial, viral and parasitic enteric pathogens (Mohapatra et al., 2007). However, current *E. coli* enumeration standard methods do not provide information on potential sources of faecal pollution. Determining the sources of origin of faecal contamination is commonly referred to as bacterial source tracking (BST) and is commonly used to assess the degree of public health risk and to ensure the development of specific approaches to reduce faecal contamination and the associated pathogens responsible for water- borne disease transmission (U. S. Environmental Protection Agency, 2002).

Several methods of bacterial source tracking (BST) have been developed in which sources of microbial contamination are identified according to differences in the characteristics of bacteria from different animal faeces. Hence, an increased understanding of the genetic variability of populations in animal reservoirs and rainwater tanks can inform epidemiological studies. Several genotypic and phenotypic BST methods have been developed in recent years to identify sources of faecal pollution (Carson et al., 2001; Mohapatra et al., 2007). Probable source(s) of faecal pollution are identified by comparing the fingerprints (phenotypic or genotypic profiles) of the environmental *E. coli* isolates with a reference library consisting of the fingerprints of *E. coli* obtained from known sources of faecal pollution.

Genotypic BST methods consider the host-specific genetic differences of indicator organisms for example ribotyping (Carson et al., 2001; Simpson et al., 2002; Scott et al., 2002); Pulsed field gel electrophoresis (PFGE) (Simpson et al., 2002); randomly amplified polymorphic DNA (RAPD) and repetitive palindromic DNA sequences (Rep PCR) (Dombek et al., 2000). Phenotypic methods employ host-specific biochemical properties, for example antibiotic resistance profiling (Hager, 2001) and carbon source utilisation tests (Harwood et al., 2003).

Antibiotic resistance in *E. coli* has been globally identified in isolates from environmental, animal and human sources. The emergence of bacterial strains resistant to antimicrobial agents is thought to have resulted from the widespread use of antibiotics for therapeutic purposes in animals and humans to control bacterial infections and may be incorporated into commercial livestock, poultry and fisheries feed at sub-therapeutic doses for growth promotion (McEwen and Fedorka-Cray, 2002). However, using antibiotics as growth promoters is believed to be the leading factor in the enhancement of antibiotic resistant bacteria selection more than the therapeutic use of antimicrobial agents in response to clinical disease (van den Bogaard et al., 2001), and may contribute to antibiotic resistance in humans acquired through food and water (Barton, 1998; Witte, 1998). To minimise this problem, banning antibiotics needed for human treatment as feed additives has been suggested, although the contribution of antibiotics as feed additives to the development of resistance in human bacterial pathogens is disputed (Prescott et al., 2000; Casewell et al., 2003).

Bacteria in the digestive tracts of humans and animals develop antibiotic resistance when they are subjected to different types, concentrations and frequencies of antimicrobial agents. Over time the exertion of positive selective pressure on bacteria-carrying genotypes conferring resistance to the antimicrobial agents leads to their proliferation (Prescott et al., 2000). Antibiotic resistance genes can be obtained by bacteria through mobile elements, such as transposons, plasmids, and integrons (Rubens et al., 1979; Prescott et al., 2000), which leads to mutations in genes responsible for antimicrobial agent uptake, binding sites (Spratt, 1994) or activation of portions of bacterial chromosomes (Hachler et al., 1991; Alekshun and Levy, 1999). The mechanisms of resistance include enzymatic inactivation or modification of antimicrobial agents, impermeability of the bacteria cell wall or membrane, active expulsion of the drug by the cell efflux pump, and alteration in target receptors (Levy, 1992). Once antibiotic resistance has been established, the genetic determinants conferring resistance can be transferred between bacteria through conjugation, transformation, and transduction (Richmond, 1972).

Concerns with the emergence of antibiotic resistant bacteria have been raised over the contamination of surface waters from livestock operations and human septage. Resistant bacteria have been isolated from a variety of sources, including domestic sewage, drinking water, rivers, and lakes (Antai, 1987; Meays et al., 2004; Donovan et al., 2008; Wose-Kinge et al., 2010). Faecal strains of *E. coli* resistant to antibiotics have been found at various prevalence levels in wild bird populations. In particular, bird populations common to areas inhabited by people and areas with a high density of livestock have been reported to be colonised with antibiotic resistant *E. coli* strains possibly selected for by the antibiotic practices in humans and domestic animals. Antibiotic resistant *E. coli* isolates have been found in corvids (Livermore et al., 2001; Literak et al., 2007; Blanco et al., 2009), house sparrows (Dolejska et al., 2008), house martins (Rybarikova et al., 2010), pigeons and doves, ducks, geese, and swans (Fallacara et al., 2001; Cole et al., 2005) and cormorants (Rose et al., 2009).

Urban birds especially pigeons and doves coexist with humans in urban environments and are considered to be important reservoirs and vectors of pathogenic and antibiotic resistant bacterial strains (Literak et al., 2007; Dolejska et al., 2009). The adaptation of these birds to urban environments is mainly ascribed to (i) the architecture of urban constructions which have gaps, cracks and spaces that can be used for landing, nesting and shelter, protecting the birds from the weather; (ii) the absence of natural predators; and (iii) availability of food (Silva et al., 2009). Animal feed lots including those for dairy cows and chicken feed waste provide an abundant source of food to the birds which then multiply extensively and become a significant source of faecal pollution. The transmission of infectious diseases or antibiotic resistant strains usually happens by either ingestion of food, water, handling of products contaminated with their faeces and dispersion of wind-dried pigeon droppings (Marques et al., 2007; Oliveira and Pinhata, 2008).

Although there have been several researches on surface and ground water contamination by antibiotic resistant bacteria, literature on antibiotic resistant bacteria in roof harvested rainwater is scarce (Antai, 1987; Meays et al., 2004; Donovan et al., 2008; Kinge et al., 2010). In rainwater harvesting, urban birds are a major source of faecal contamination as they have direct access to the roof catchment surface and have been implicated as reservoirs and vectors for the spread of antibiotic resistant strains of *E. coli* (Fallacara et al., 2001; Cole et al., 2005; Dolejska et al., 2008). Resistance to more than one antibiotic by a single bacterial isolate is commonly reported. Consequently, resistance to multiple antibiotics have been used to identify and differentiate *E. coli* strains from different animal species (Dolejska et al., 2008). The use of antibiotic resistance profiles is simple, cost-effective, and suitable for surveillance (Lewis et al., 2007), and has been used for *E. coli* strains to identify sources of faecal contamination in water (Clermont et al., 2000; Dobbin et al., 2005; Costa et al., 2006; Clinical and Laboratory Standards Institute (CLSI), 2008).

The use of antimicrobial agent resistance profiles to identify sources of bacterial contamination is a promising and emerging procedure. Cluster analysis technique of antimicrobial agent resistance profiles has been reported to be a useful, low-cost screening method (Livermore et al., 2001). It can be used in both phenotypic and genotypic BST to classify bacteria into their respective similarity groups. Cluster analysis is a statistical

technique that uses a number of variables (multivariate) to classify/group subjects into clusters in a dendrogram tree (Stevens, 1996). The analysis sequentially converts similarity data into more inclusive groupings, combining like strains into clusters based on their similarity coefficient. Two very similar strains are grouped together and then joined with another cluster to form a new, large and more inclusive cluster. This is repeated until all strains and associated clusters are tied completely together. Several mathematical clustering methods exist, including the unweighted pair groups' method analysis (UPGMA), weighted pair groups method analysis (WPGMA), Ward, Complete Linkage and Single Linkage (Romesburg, 1990).

If the use of antibiotics is an important factor for the development of antimicrobial agent resistance, as suggested from current research, it therefore can be hypothesised that antibiotic resistance patterns in different animal and environmental populations vary according to the extent of exposure including the antibiotic type and quantities. We here evaluated this assumption on the impact of wild birds, on roof harvested rainwater from five sites representing two different areas. This was accomplished first by identifying antibiotic resistance patterns of *E. coli* strains obtained from wild birds and comparing them to those of *E. coli* strains obtained from roof harvested rainwater.

## **6.2 Methodology**

### **6.2.1 Area of the study and collection of samples**

Roof harvested rainwater samples were collected from five sampling sites which included, three in Pretoria (PTA); the Experimental Farm, University of Pretoria Experimental Farm (PTA Site 1(S1)), the Plant Science Building, University of Pretoria Hatfield campus (PTA Site 2 (S2) and a house in Sunnyside suburb (PTA Site 3 (S3), and two sites in Johannesburg (JHB); Endluweni Primary School (JHB S1) and Thembisa Secondary School (JHB S2). Faecal samples were collected from urban pigeons (*Columbia livia*) at the University of Pretoria Experimental Farm only. This site (PTA S1) represents a typical farm setting; with cattle feed pen within 100m, where a large number of pigeons come to feed. These pigeons also feed on chicken layers waste from fowl runs approximate 800m away. The house where three rainwater tanks are installed (PTA S1) has overhanging Mulberry trees on the roof,

where various kinds of birds feed on mulberries. The Plant Science Building site (PTA S2) is located at the University of Pretoria Main Campus. This site is on the second floor and has three tanks installed. The site represents a typical urban scenario where there is minimal vegetation and bird interference. The PTA S3 is located in Sunnyside suburb in Pretoria. This site is surrounded by trees where various kinds of birds nest. The Johannesburg sites (JHB S1 and JHB S2) are schools (each with two rainwater tanks) located in Thembisa, a high density suburb with limited number of trees and various birds which feed on garbage dumps.

## **6.2.2 Sample collection and isolation of presumptive *Escherichia coli* strains**

### **6.2.2.1 Rainwater Sample collection**

Samples were collected from eleven rainwater tanks during the 2012 – 2013 rain seasons, from September to February within 1 to 4 days after a rain event (ranging from 35 to 130 mm) as previously described by (Ahmed, Huygens, Goonetilleke and Gardner, 2008). In total 74 RHRW samples (of which forty-four tested positive to *E. coli*) were collected in duplicates from the outlet taps located close to the base of the tanks, in sterilised two litre containers. Taps were wiped with 70% ethanol, and allowed to run for 30 to 60 seconds to flush out stagnant water from the taps before collecting water samples. Collected samples were transported to the laboratory and processed within 10 hours.

Undiluted RHRW samples were assayed directly for densities of faecal coliforms and *E. coli*, with Colilert chromogenic substrate tests kits and Quantitray 2000 MPN trays (Idexx, Westbrook, Maine) as per the manufacturer's instructions. The inoculated Quantitrays were subsequently sealed and incubated at 35°C for 24-28h. Following incubation, the Quantitray wells were read for yellow colour and fluorescence. A bench top ultraviolet (UV) light (366 nm) was used to identify fluorescent wells. A manufacturer provided MPN table (Idexx, Westbrook, Maine) was used to generate microbial density estimates based on the proportion of positive reactions in each tray.

### **6.2.3 Faecal sample collection**

Fresh faeces from thirty birds including pigeons and doves were collected. Indicator bacterial density in faecal material was measured by first diluting one gram of faeces in 9ml of

distilled water, vortexed and allowed to stand for 5 minutes to allow debris to settle. A 1ml sample of the supernatant was consequently extracted and serially diluted for microbial isolation and quantification. Densities of faecal coliforms and *E. coli*, and enterococci were determined with Colilert-18 as described above.

#### **6.2.4 Recovery**

Following incubation, the backing material of each Quantitray was disinfected by application of 70% ethanol with a sterile swab. After the residual ethanol had evaporated, sterile razor blades were used to pierce the backing material of three fluorescence positive wells per tray and three trays were processed per water sample. One loop full of well content was streaked onto mEndo selective agar for the isolation of *E. coli*. Presumptive *E. coli* colonies were sub-cultured twice on nutrient agar and re-inoculated into the respective Collilert chromogenic media in sterile 96 micro well plates and incubated for 18 hours to confirm fluorescence before MALDI-TOF-MS and PCR analysis.

#### **6.2.5 MALDI-TOF-MS identification and characterisation of bacterial isolates**

Bacterial strains were sub-cultured twice on nutrient agar (Merck, Johannesburg) before MALDI-TOF-MS analysis. After 24 to 48-hour cultivation of an isolate on nutrient agar, a single colony was transferred with a toothpick onto MALDI plates in duplicates (Sigma-Aldrich, USA). The preparation was overlaid with 1  $\mu\text{l}$  of saturated solution of cyano-4-hydroxycinnamic acid in organic solution (50% acetonitrile, 2.5% trifluoroacetic acid), crystallised by air-drying at room temperature and directly screened (Bittar et al., 2009; Pinto et al., 2011). Mass spectra were generated with the Microflex LT mass spectrometer operated by the MALDI Biotyper automation control and recorded by Flex Control software (Bruker Daltonics, Bremen, Germany). Three hundred shots per sample spot were acquired using the recommended instrument settings for bacterial identification (linear positive mode, 60 Hz laser frequency, 20 kV acceleration voltage, 16.7 kV IS2 voltage, 170 ns extraction delay, and 2,000 to 20,137 m/z range). The peak lists generated was used for matches against the reference library directly using the integrated pattern matching algorithm of the software. The whole process from MALDI-TOF-MS measurement for identification was performed automatically without user intervention.



MALDI Biotyper 3.0 software (Bruker Daltonics) was used to analyse raw spectra of the bacterial isolates with default settings. The software compares acquired sample spectra to reference spectra in the provided database and compiles a list of best matching database records. The degree of spectral pattern matching is expressed as a logarithmic identification score and interpreted according to the manufacturer's instructions. Results are expressed as log (score) values ranging from 0 to 3 levels. Scores  $\geq 2.300$  indicate species identification with a high level of confidence,  $\geq 2.000$  indicate species identification,  $< 1.700$ – $1.999$  indicate genus identification, and  $> 1.700$ , no identification (Romanus et al., 2011).

### **6.2.6 Polymerase chain reaction for detection of UidA gene in *Escherichia coli***

The polymerase chain reaction (PCR) was used to detect the presence of the UidA gene, which codes for the  $\beta$ -D-glucuronidase enzyme. A 147 bp coding region of the *E. coli* uid gene was amplified using the 20 and 21-mer primers UAL-754 (5'-AAAACGGCAAGAAAA-AGCAG-3') and UAR-900 (5'-ACGCGTGGTTACAGTCTTGCG-3') (Bej et al., 1991). These primers were synthesised by Inqaba Biotech (Pretoria, South Africa). An ICycler thermal cycler (Bio-Rad, UK) was used to amplify the DNA. In both PCRs the final reaction volume of 25  $\mu$ l consisted of 12.5  $\mu$ l double strength PCR master mix (0.05U/ $\mu$ l *Taq* DNA Polymerase in reaction buffer, 0.4mM of each dNTP (dATP, dCTP, dGTP, dTTP), 4mM MgCl<sub>2</sub>; (Fermentas Life Science, US), PCR-grade water (Fermentas Life Science, US), 50ng sample DNA and 25pmole of each primer. Amplification was performed with a thermal cycler programmed for 1 cycle of 2 min at 94°C; 25 cycles of 1min at 94°C, 1 min at 58°C, 2 min at 72°C; 1 cycle of 5min at 72°C. PCR products were electrophoresed on a 1.5% agarose gel, stained with GR Green (Fermentas Life Science, USA), and visualised using Gel Documentation System (Bio-Rad Gel Doc EZ Imager, USA).

### **6.2.7 Antimicrobial susceptibility testing**

The Kirby-Bauer disk diffusion technique (Bauer et al., 1966) was used to determine antibiotic susceptibility profiles of 239 *E. coli* isolates. Isolates were cultured at 37°C for 24 hours in nutrient broth (Merck, Johannesburg, South Africa) and the suspension was adjusted

to a turbidity equivalent to a 0.5 McFarland standard before spreading 100µl onto Mueller Hinton agar plate (Merck, Johannesburg, South Africa). The disks used (Mast Diagnostics, UK, supplied by Davies Diagnostics, Midrand, SA) included ampicillin (Ap, 10 µg), amoxicillin (A, µg10), amikacin (Ak, 30µg), cefoxitin (Fox, 30 µg) ceftriaxone (Cro, 30 µg), chloramphenicol (C, 10µg), ciprofloxacin (Cip, 5 µg), cotrimoxazole (TS, 25 µg), enrofloxacin (Enf, 5µg) gentamicin (Gn, 10 µg), nalidixic acid (Na, 30 µg) and tetracycline (T, 30 µg) are shown in Table 6.1. Isolates were categorised as susceptible or resistant based upon interpretive criteria developed by the CLSI, 2007).

Table 6. 1: Details of the antibiotics that were used in the study to test for antibiotic resistance

Group	Antibiotic	Abbreviation	Antibiotic disc concentrations (µg)	Inhibition zone (mm)		
				Resistant	Intermediate resistant	Susceptible
Penicillin's	Ampicillin	A	10	≤13	14-16	≥17
	Amoxicillin	AP	10	≤13	14-16	≥17
Aminoglycosides	Amikacin	AK	30	≤14	15-16	≥17
	Gentamicin	GM	10	≤12	13-14	≥15
Quinolones	Enrofloxacin	ENF	5	≤17	18-20	≥21
	Cefoxitin	FOX	30	≤14	15-17	≥18
Phenicol	Chloramphenicol	C	10	≤12	13-17	≥18
Ciprofloxacin	Ciprofloxacin	CIP	5	≤15	16-20	≥21
Cephems	Ceftriaxone	CRO	30	≤13	14-20	≥21
Quinolones	Nalidixic acid	NA	30	≤13	14-18	≥19
Folate Pathway Inhibitors	Cotrimoxazole	TS	25	≤10	11-15	≥16
Tetracyclines	Tetracycline	T	30	≤11	12-14	≥15

Source: The concentration used as well as the inhibition zone measurements were according to the National Committee on Clinical Laboratory Standards (CLSI, 2007)

Note: The abbreviations are as they appeared on the antibiotic discs.

### 6.2.8 Statistical analysis

All statistical analysis were carried out using Statistica 10 (Stat soft, US). Data for the antimicrobial agent resistance of each bacterial isolate were reported in two formats: either as the diameter of the zone of inhibition (in millimetres) or as susceptible or resistant (based on CLSI breakpoints). Since these data were used to evaluate the grouping isolates by their geographical origin (i. e. by sites within specific areas) and by their source (i. e. RHRW or pigeon faeces). RHRW isolates were handled in 2 different ways, (i) individually by site (PTA S1, PTA S2, PTA S3, JHB S1 and JHB S2), by groups based on area (PTA or JHB)

and source (RHRW). Associations between isolate source groups, RHRW areas and sites, and associated antimicrobial agent profile (resistant or not resistant) were expressed as percentages. The significance of differences in zones of inhibition between species groups was calculated with the Kruskal Wallis test (alpha value, 0.05) and the Pearson product moment correlation was used to compare the similarity relationship between the groups.

Susceptibility data on the *E. coli* isolates from the different sites and sources tested against the twelve antibiotics, were used for cluster analysis by means of Ward's method and Euclidean distances as a measure of similarity. Ward's clustering method is a hierarchical agglomerative method whose objective is to create clusters that give minimum increase in the total within group error sum of squares.

### **6.3 Results**

A total of thirty faecal samples from urban pigeons (*Columbia livia*) from Pretoria and seventy-four RHRW samples from eleven rainwater tanks in Johannesburg (4) and Pretoria (7) were collected. We detected *E. coli* from all pigeon faeces and RHRW sites, although only forty-four of the 74 RHRW samples tested were positive for *E. coli*. Generally, levels of RHRW contamination with *E. coli* were relatively higher in water samples obtained from PTA S1 and PTA S3 than the rest of the RHRW sites. Data on the prevalence and abundance of enteric microbial populations are reported in chapter 4. Table 6.3 shows the number of isolates that were used in this study and their respective sources. A total of 328 presumptive *E. coli* isolates were isolated from both pigeon droppings and RHRW samples, sub-cultured and further analysed. However, only 239 isolates satisfied all the identification criteria and were used for subsequent analysis. The 239 isolates were comprised of 130 isolates from pigeon faeces, 78 from Pretoria and thirty-one from Johannesburg. Isolates from Pretoria sites were comprised of 10, 18 and 50 isolates from PTA S3, PTA S2 and PTA S1 whereas those from Johannesburg were from JHB S2 (10) and JHB S1 (21).

#### **6.3.1 Antibiotic susceptibility**

All 239 *E. coli* isolates were subjected to antibiotic susceptibility tests with twelve antimicrobial agents, from ten different antibiotic classes (Table 6.1). Antibiotic resistance

was detected in both pigeon droppings and RHRW *E. coli* isolates (Table 6.2). The most frequently encountered form of resistance in all samples was resistance to ampicillin. In all samples, resistance to ampicillin (22.7.9%), gentamicin (23.6%), amikacin (24%), tetracycline (17.4) and amoxicillin (16.9%) were the most frequently encountered form of resistance. However, a relatively higher proportion of isolates from pigeon faeces were antibiotic resistant than those from RHRW tanks, with the most significant differences being noted in resistance to gentamicin (12.3%), tetracycline (9.8%), chloramphenicol (6.5%) and amikacin (4.2%). Pigeon droppings and RHRW *E. coli* isolates exhibited resistance to eleven of the twelve antibiotics tested, with cotrimoxazole resistance not being detected in bird samples and ceftriaxone resistance in RHRW samples. Other forms of resistance detected in small proportions were on ceftriaxone (1.5%) and nalidixic acid (3.8%) in pigeon droppings and cotrimoxazole (0.9%), nalidixic acid (0.9%) chloramphenicol (1.8%) ciprofloxacin (3.7%), ceftiofloxacin (3.7%) and ciprofloxacin (4.6%) in RHRW samples.

Table 6. 2: Antibiotic resistance profiles of *Escherichia coli* isolated from roof harvested rainwater originating from Pretoria and Johannesburg, and bird faecal samples in Gauteng Province of South Africa

Antibiotic	Symbol	RHRW Sources										RHRW <sup>a</sup>		Bird		Totals	
		JHB		JHBS2		PTAS3		PTAS1		PTA		n	%	n	%	n	%
		n	%	n	%	n	%	n	%	n	%						
Amoxicillin	A	2	9.5	-	-	2	20	13	26	3	16.7	20	18.3	21	15.8	41	16.9
Amikacin	AK	9	42.9	1	10	-	-	10	20	4	22.2	24	22	36	27.1	60	24.8
Ampicillin	AP	1	4.8	2	20	1	10	11	22	5	27.8	20	18.3	35	26.3	55	22.7
Chloramphenicol	C	1	4.8	-	-	-	-	-	-	1	5.6	2	1.8	12	9	14	5.8
Ciprofloxacin	CIP	1	4.8	-	-	1	10	1	2	-	-	3	2.8	10	7.5	13	5.4
Ceftriaxone	CRO	-	-	-	-	-	-	-	-	-	-	-	-	2	1.5	2	0.8
Enrofloxacin	ENF	4	19	-	-	-	-	-	-	-	-	4	3.7	8	6	12	5
Ceftiofloxacin	FOX	1	4.8	-	-	-	-	3	6	-	-	4	3.7	7	5.3	11	4.5
Gentamicin	GM	6	28.6	3	30	-	-	10	20	2	11.1	21	19.3	36	27.1	57	23.6
Nalidixic Acid	NA	-	-	-	-	-	-	1	2	-	-	1	0.9	6	4.5	7	2.9
Tetracycline	T	6	28.6	1	10	-	-	2	4	4	22.2	13	11.9	29	21.8	42	17.4
Cotrimoxazole	TS	1	4.8	-	-	-	-	-	-	-	-	1	0.9	3	2.3	4	1.7

Note: Percentages were calculated by dividing the number of *E. coli* isolates confirmed as antibiotic resistant in a particular sample site by the total number of isolates tested for the particular site or group.

<sup>a</sup>All rooftop harvested rainwater sites (RHWR)

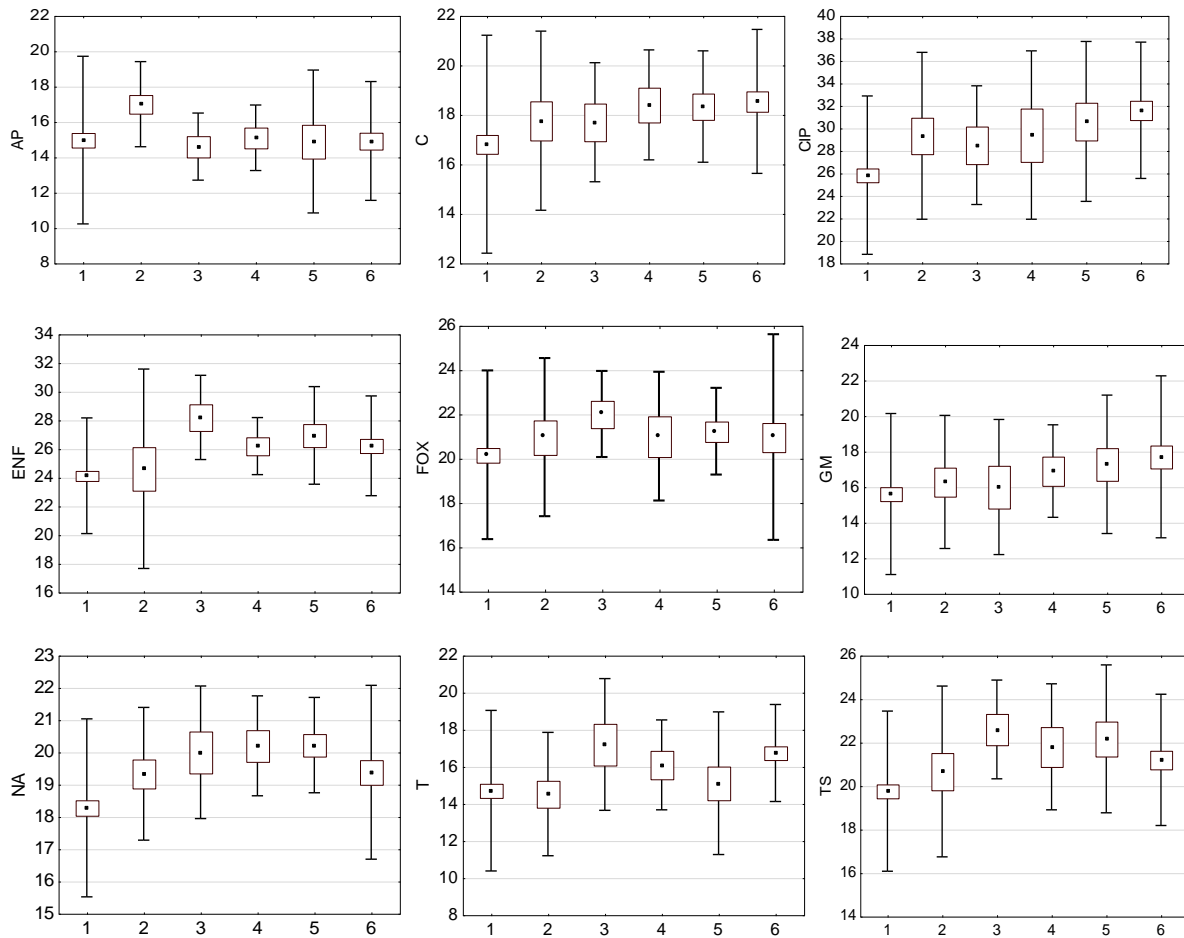
### 6.3.2 Non-parametric test

Kruskal Wallis test was performed on original disk diffusion zone sizes data with respect to the different sources for each antibiotic tested (Table 6.3). The results showed that antibiotic resistance/susceptibility patterns were significantly different for nine of the twelve antibiotics tested, with no significant differences being detected for the three antibiotics; amoxicillin, ampicillin and ceftriaxone. Since the Kruskal Wallis test leads to significant results, when at least one of the samples is different from the other samples but does not identify where the differences occur or how many differences actually occur. We further compared rainwater sample sources only and significant differences ( $p \leq 0.05$ ) were observed on resistance to ampicillin only. This implies that the source of significant differences were pigeon faecal samples for 8 of the nine antibiotics. Consequently, the largest inhibition zones (indicating greater susceptibility) were found in RHRW sources for eight of the antibiotics on which significant differences had been observed (Figure 6.1). When resistance to ampicillin among RHRW sites was considered significantly larger inhibition zones were found with JHB S1 compared to the rest of the other sources (Table 6.3).

Table 6. 3: Kruskal Wallis test of differences between pigeon and roof harvested rainwater isolate sources

Antibiotics	Mean disk diffusion zone diameter (mm)						Kruskal Wallis ( <i>P-value</i> )	
	Bird (n=130)	JHB S1 (n=21)	JHB S2 (n=10)	PTA S3 (n=10)	PTA S2 (n=18)	PTA S1 (n=50)	All Sources	RHRW <sup>b</sup>
Amoxicillin	15.2	16.7	17.5	14.8	15.6	14.8	.115	.196
Amikacin	16.9	17.2	17.9	18.9	18.1	18.3	.249	.717
Ampicillin	15	17	14.6	15.1	14.9	14.9	.054	.041
Chloramphenicol	16.8	17.8	17.7	18.4	18.3	18.5	.000	.895
Ciprofloxacin	25.8	29.3	28.5	29.4	30.6	31.6	.000	.466
Ceftriaxone	25	25.1	26.2	25.9	25.9	25.5	.789	.928
Enrofloxacin	24.1	24.6	28.2	26.2	26.9	26.2	.001	.464
Cefoxitin	20.2	21	22	21	21.2	21	.006	.921
Gentamicin	15.6	16.3	16	16.9	17.3	17.7	.037	.653
Nalidixic Acid	18.3	19.3	20	20.2	20.2	19.4	.002	.397
Tetracycline	14.7	14.5	17.2	16.1	15.1	16.7	.053	.127
Cotrimoxazole	19.8	20.7	22.6	21.8	22.2	21.2	.001	.694

Significance level at  $p \leq 0.05$ , <sup>a</sup> tests among all isolate sources, <sup>b</sup> tests among RHRW sources only.



Note: The vertical axis represents the zone diameter of inhibition (mm) and the horizontal axis represents the different sources of isolates as follows, pigeon faecal isolates (1), RHRW isolates from JHB S1 (2), JHB S2(3), PTA S3 (4) PTA S2 (5) and PTA S1 (6). The antibiotics represented are ampicillin (Ap), chloramphenicol (C), ciprofloxacin (Cip), enrofloxacin (Enf), ceftiofur (Fox), gentamicin (Gm), nalidixic acid (Na) tetracycline (T) and cotrimoxazole (Ts).

Figure 6. 1: The variability in antibiotic resistance/susceptibility patterns for where significant differences observed between the different sources: each box plot represents the mean (dot) +/- standard error (bottom and top of box) +/- standard deviation (whiskers)

### 6.3.3 Correlation test

Table 6.4 shows the comparison of percentages of antibiotic resistant *E. coli* isolated from RHRW sources in Johannesburg and Pretoria, and from pigeon faeces. No significant correlation was observed between Johannesburg and Pretoria sources ( $r \leq 0.611$ ,  $p < 0.01$ ), although there was a relatively high correlation between source sites from the same area ( $r \geq 0.707$ ), except for isolates from PTA S3 ( $r=0.6$ ). This PTA S3 site had no significant correlation to any other site. However, all other RHRW sites were significantly correlated to

the averages from RHRW and bird sources ( $r \geq 7.17$ ) except for the JHB S1 ( $r=0.684$ ) source with RHRW source. Antibiotic resistance profiles of pigeon isolates had a wide diversity representation, with the result that, relatively high correlation rates existed between bird and individual RHRW sites. However, individual RHRW sites had relatively limited diversity unique to individual sites, resulting in lower correlation rates between sites.

Table 6. 4: Pearson's product moment correlation coefficient between different sites, areas and sources of *Escherichia coli* antibiotic resistance profiles

	JHB S1	JHB S2	JHB	PTA S1	PTA S2	PTA S3	PTA	RHRW Total
JHB S1	1							
JHB S2	0.514	1						
JHB	0.970**	0.707*	1					
PTA S1	0.41	0.611*	0.511	1				
PTA S2	0.541	0.611*	0.618*	0.748**	1			
PTA S3	-0.225	-0.047	-0.198	0.573	0.339	1		
PTA	0.425	0.606*	0.522	0.983**	0.848**	0.600*	1	
RHRW Total	0.684*	0.721**	0.768**	0.930**	0.868**	0.376	0.947**	1
Pigeon faeces	0.717**	0.835**	0.827**	0.792**	0.899**	0.193	0.838**	0.940**

Significance of t-test; \* 0.05 and \*\* 0.01 level (1-tailed).

### 6.3.4 Antibiotic resistance phenotypes

Antibiotic resistance (AR) phenotypes were determined for *E. coli* isolated from RHRW and faeces obtained from storage tanks and urban pigeons, respectively (Table 6.4). Figure 6.2 shows the observed antibiotic resistant percentages of isolates by their sources for 1-5 and 7 antibiotics. The highest number of phenotypes was observed for single antibiotics (7), whereas 4, 7, 5, 1 and 1 multiple antibiotic resistance (MAR) phenotypes were detected for combinations of 2, 3, 4, 5 and 7 antibiotics, respectively. No single AR phenotypes were detected for chloramphenicol, ceftriaxone, gentamicin, cefoxitin and cotrimoxazole, although, they were detected in MAR phenotypes. The most abundant resistance phenotype was on single antibiotics with 34.4, 20.5 and 22.6% of isolates from JHB, PTA and pigeon faeces being resistant. For these resistance percentages tetracycline (6.2%) and ampicillin

(5.8%) were the most abundant, considering all isolates. A higher proportion of isolates from JHB (18.8%) had tetracycline resistance phenotype compared to those from Pretoria (3.8%) and pigeon faeces (4.5%). However, isolates of relatively the same proportion (5.3-6.4%) from all sources were resistant to ampicillin. The least abundant single antibiotic resistance phenotype (1.2%) was observed for ciprofloxacin and nalidixic acid on isolates from Pretoria and pigeon faeces.

Table 6. 5: Antibiotic resistance phenotypes of *Escherichia coli* isolated from pigeon faeces and roof harvested rainwater obtained from rainwater harvesting tanks\*

Resistance Phenotype	Source								Overall	
	Pretoria		Johannesburg		RHRW		Birds			
	n	%	n	%	n	%	n	%	n	%
A	3	3.8	1	3.1	4	3.6	1	0.8	5	2.1
Ak	2	2.6	1	3.1	3	2.7	6	4.5	9	3.7
Ap	5	6.4	2	6.3	7	6.4	7	5.3	14	5.8
Cip	1	1.3	0	0	1	0.9	2	1.5	3	1.2
Gm	1	1.3	1	3.1	2	1.8	6	4.5	8	3.3
Na	1	1.3	0	0	1	0.9	2	1.5	3	1.2
T	3	3.8	6	18.8	9	8.2	6	4.5	15	6.2
A-Fox	2	2.6	0	0	2	1.8	1	0.8	3	1.2
Gm-Ak	1	1.3	2	6.3	3	2.7	10	7.6	13	5.4
Gm-Ap	2	2.6	1	3.1	3	2.7	2	1.5	5	2.1
Ak-Ap	2	2.6	0	0	2	1.8	2	1.5	4	1.7
A-C-T	0	0	0	0	0	0	6	4.5	6	2.5
A-Ap-Fox	2	2.6	0	0	2	1.8	3	2.3	5	2.1
A-Ap-T	2	2.6	0	0	2	1.8	2	1.5	4	1.7
A-Ap-TC	1	1.3	0	0	1	0.9	3	2.3	4	1.7
A-Ak-Gm	3	3.8	0	0	3	2.7	2	1.5	5	2.1
Ak-Ap-Cip	1	1.3	1	3.1	2	1.8	4	3	6	2.5
Ak-Ap-Gm	2	2.6	0	0	2	1.8	2	1.5	4	1.7
Ak- Gm-Enf- Ap	0	0	2	6.3	2	1.8	1	0.8	3	1.2
Ak- Gm-Enf- Cro	0	0	2	6.3	2	1.8	1	0.8	3	1.2
Ak-Gm-Ap-A	3	3.8	1	3.1	4	3.6	0	0	4	1.7
Ak-Gm-Cip-T	1	1.3	1	3.1	2	1.8	9	6.8	11	4.5
Gm-Cip-Fox-A	0	0	0	0	0	0	5	3.8	5	2.1
Ak-Enf-Gm-Ap-T	0	0	0	0	0	0	4	3	4	1.7
A-Cip-Gm-Ap-Cro-Fox-Na	0	0	0	0	0	0	2	1.5	2	0.8
Susceptible	40	51.3	11	34.4	51	46.4	43	32.6	94	38.8

Note: Sampling sites were grouped by their origin to form the Pretoria and Johannesburg areas which were also combined to form the roof harvested rainwater source. Percentages were calculated by dividing the number of confirmed antibiotic resistant *E. coli* in a particular area or source by the total number of isolates tested for the particular area or source.

Resistance to two antibiotics was relatively similar (9.1-9.4%) among RHRW sources, although bird isolates had a relatively higher proportion (15.9%) resistant to the same number



of antibiotics. The A-Fox and Ak-Ap combinations were not detected in isolates from Johannesburg, although they were present in isolates from Pretoria and pigeon faeces. The most dominant phenotype of resistance to two antibiotics was Ak-Gm (5.4%). A relatively higher proportion of isolates with this phenotype was detected in pigeon faeces (7.6%) and Johannesburg area. A relatively higher proportion of isolates from Pretoria (14.2%) and pigeon faeces (12.1%) had MAR for combinations of three antibiotics compared to those from Johannesburg (3.1%). The most abundant three antibiotic phenotype was A-Ak-Gm with prevalence's ranging from 3.1- 3.8% for all sources. Of the six MAR phenotypes to combinations of three antibiotics, only 1 (Ak-Ap-Cip) was detected from Johannesburg isolates, whereas all the six phenotypes were detected in isolates from Pretoria and pigeon faeces. Furthermore the A-C-T phenotype was detected in isolates from pigeon faeces (4.5%) only.

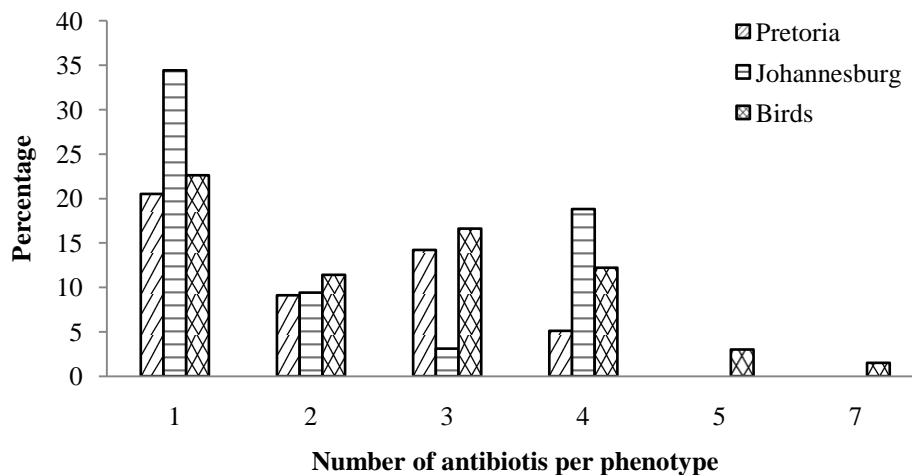


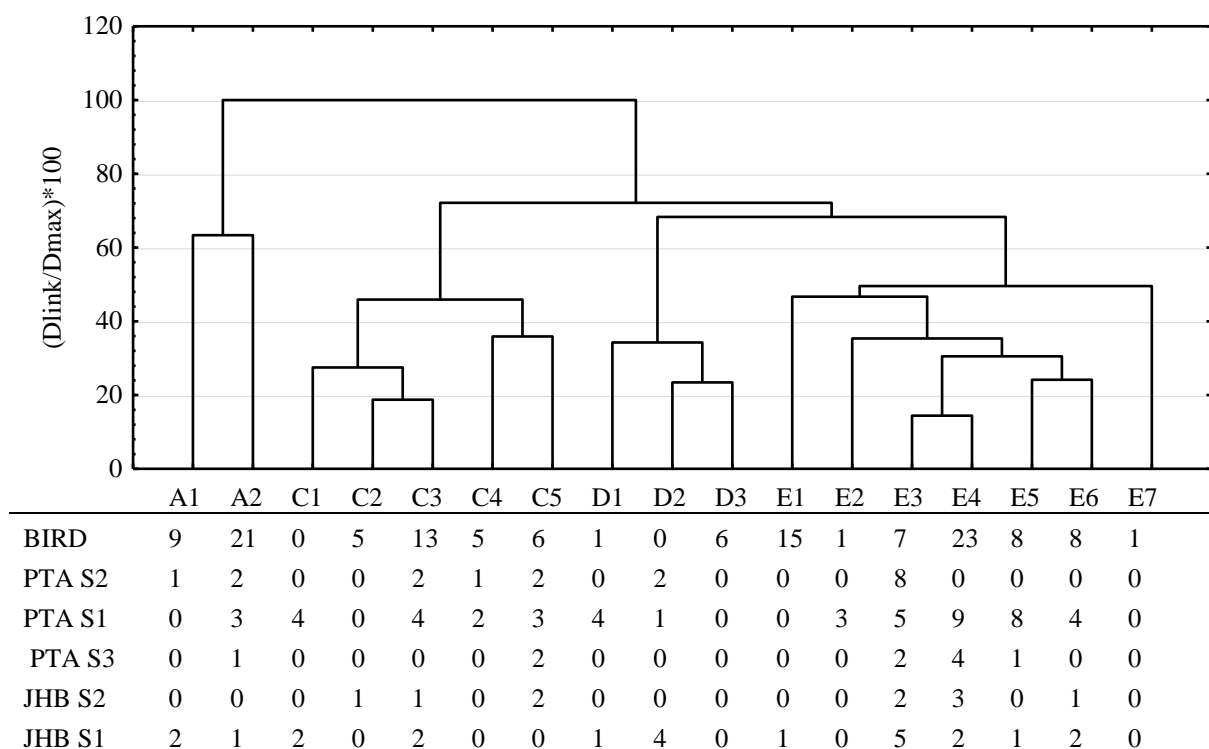
Figure 6. 2: Abundance percentages of the observed phenotypes per groups of antibiotics by their sources

The highest MAR observed for combinations of four antibiotics was for isolates from Johannesburg (18.8%) and pigeon faeces 15.2%, although only 5.1% of Pretoria isolates were resistant to the same number of antibiotics. The most abundant resistance phenotype to four antibiotics (4.5%) was Ak-Gm-Cip-T. This phenotype was dominated by isolates from birds (6.8%) with Pretoria and Johannesburg isolates having low proportions of 1.3% and 3.1% respectively. Similar phenotypes were observe for the Ak-Enf-Gm-Ap and Ak-Enf-Gm-Cro antibiotic combinations for isolates from Johannesburg, (6.3%) but were not detected in isolates from Pretoria. The four (Gm-Cip-Fox-A), 5 (Ak-Enf-Gm-Ap-T) and seven (A-Cip-

Gm-Ap-Cro-Fox-Na) were detected in four, five and two isolates from birds only, respectively. However, the phenotypes Ak-Enf-Gm-Ap and Ak-Enf-Gm-Cro were not detected from Pretoria isolates, whereas Ak-Gm-Ap-A was only detected in RHRW sources.

### 6.3.5 Cluster analysis

Cluster analysis was applied to the 242 *E. coli* isolates from RHRW and bird sources using the antibiotic inhibition zone diameter data. This was done to analyse the relationship between isolates by their sources and the possible link with pigeon faeces as the major source of contamination. The dendrogram was constructed using the Ward method and Euclidean linkage distances. The linkage distance is reported as  $(Dlink/Dmax)$ , which represents the quotient between the linkage distance for a particular case divided by the maximal distance, multiplied by 100 as a way to standardise the linkage distance represented on Y-axis (Wunderlin et al., 2001; Singh et al., 2004).



Note: Bacterial Source designation prefixes are based on the source (bird or RHRW) and area (Johannesburg (JHB) or Pretoria (PTA) and the site number for the individual areas (as described in the methodology section).

Figure. 6. 3: Dendrogram showing the relationship between *Escherichia coli* isolated from pigeon faeces and roof harvested rainwater samples obtained from sites in Johannesburg and Pretoria.

Cluster analysis yielded a dendrogram (Figure 6.3), grouping the isolates into five (A, B, C D and E) statistically significant clusters at  $(Dlink/Dmax) \times 100 < 65$ . At  $(Dlink/Dmax) \times 100 < 10$ , the five major clusters subdivided into seventeen minor clusters (A1-E7) as shown in Figure 6.3. Further analysis of the clusters was performed for patterns of associations between isolates from the various sites, areas and sources. The analysis obtained was used as a tool in determining the uniqueness between antibiotic resistance patterns of *E. coli* isolates from different sources. Clusters A1, C2, D3, E1 and E7 were relatively unique to pigeon faeces, whereas clusters C1, D1, D2 and E2 were relatively unique to RHRW sources. Of the twelve isolates in cluster A1, 9 were from pigeon faeces and only one and two isolates were from PTA S2 and JHB S1 sites, respectively. Similarly all 6 and 14 of 15 isolates in Clusters D3 and E1 originated from pigeon faeces, respectively. For RHRW samples, no pigeon faecal isolates were detected in clusters C1 and D2. Cluster C1 had 4 and 2 isolates from PTA S1 and JHB S1, respectively. Cluster D2 had seven isolates from PTA S1 (2), PTA S2 (1) and JHB S1 (4). However when RHRW sources were considered, clusters A2, C4, D1 E2 and E5 were predominantly comprised of isolates from Pretoria, whereas clusters C2 and E1 were comprised of JHB isolates. Of the seven RHRW isolates in cluster A2 only one was from the JHB area and all RHRW isolates in cluster E2 were from PTA S1. Similarly, the only RHRW isolate in E1 was from JHB S1. The clusters A2, C3, C5, E3, E4 and E6 showed *E. coli* isolates from all sources. Clusters C5 and E3 had at least two representative isolates from all sources, except for JHB S1 in cluster C5 for which there were no representative isolates. Although a significant number of *E. coli* isolates from all the sources shared a number of common clusters, some clusters were unique to RHRW or pigeon faeces, or to specific RHRW sampling areas or sites.

#### 6.4 Discussion

South Africa has a mix of developed and developing regions with at least 9.7 million (20%) of the people not having access to adequate water supply (Kahinda et al., 2007). RHRW appears to be one of the most promising alternatives for supplying freshwater in the face of increasing water scarcity and escalating demand (Sazakli et al., 2007). Rainwater harvesting (RHRH) can provide water directly to households including those in rural and peri-urban areas where conventional technologies cannot supply. In order to improve water access, the

South African government has committed itself to provide RHRW tanks to poor households, who in most instances use it without prior treatment as it is assumed to be safe (Kahinda et al., 2007). However, RHRW is known to be prone to contamination with enteric pathogens including *E. coli* (Ahmed et al., 2008; 2010b; 2011b; Nawaz et al., 2012; Kaushik et al., 2012). Given the South African context, the need for good domestic water quality takes another dimension, when one considers that the country has one of the worst HIV/AIDS epidemic in the world, with the number of infected people estimated at 5.5 million (UNAIDS, 2006).

To help evaluate the risk associated with the use of RHRW and the significance of pigeon faeces as the most likely major source of contamination, we characterised antibiotic resistance profiles in *E. coli* isolated from RHRW tanks and urban pigeon faeces. A total of 239 *E. coli* were tested for their susceptibility trends against a panel of twelve common antimicrobial agents. The isolates originated from faeces of urban pigeons (*Columbia livia*), and RHRW obtained from Johannesburg and Pretoria in Gauteng Province of South Africa. Sampling sites could be categorised into farm and urban settings, and urban settings could be divided into low and high density suburbs. The objective of this study was to evaluate the prevalence of antibiotic resistance and compare antibiotic resistance profiles of *E. coli* isolated from pigeon faeces and RHRW from different environmental settings.

The motivation for this was the fact that previous studies on RHRW have implicated droppings of birds, mammals and/or rodents that have access to catchment areas or water storage tanks as the possible sources of microbial contamination (Ahmed et al., 2008; Amin and Ham, 2009; Martin et al., 2010). Of these animals birds are the most mobile and have been implicated as reservoirs of antibiotic resistant enteric pathogens. Several investigations into RHRW contamination detected *E. coli* as the most common indicator of faecal contamination (Pinfold et al., 1993; Sazakli et al., 2007; Ahmed et al., 2008; Amin and Ham, 2009; Martin et al., 2010). Hence we selected *E. coli* as our indicator bacteria.

Bird populations sympatric to areas inhabited by people and areas with a high density of livestock have been reported to be colonised with antibiotic resistant *E. coli* strains possibly

selected by the antibiotic practice in humans and domestic animals (Livermore et al., 2001; Fallacara et al., 2001; Cole et al., 2005; Literak et al., 2007; Dolejska et al., 2008; Blanco et al., 2009; Rose et al., 2009; Rybarikova et al., 2010). In particular, urban pigeons (*Columba livia*) come into close contact with humans and other animals, and are considered to be important reservoirs and vectors of pathogenic and antibiotic resistant bacterial strains (Kobayashi et al., 2002; Haag-Wackernagel and Moch, 2004; Wani et al., 2004; Literak et al., 2007; Dolejska et al., 2009). The adaptation of these birds to urban environments, the absence of natural predators and availability of food enables them to multiply extensively and become a significant source of faecal pollution (Silva et al., 2009). The presence of pigeon faeces in urban environments may contribute to the spread of infectious agents through the dispersion of wind-dried dropping particles (Marques et al., 2007; Oliveira and Pinhata, 2008).

We therefore hypothesised that if birds are the major source of faecal pollution, then the antibiotic resistance profiles between RHRW isolates from PTA S1 and pigeon faecal isolates would be similar. Since the birds are mobile, we included a site (PTA S2) 5 km away, that had limited bird interference and another (PTA S3) that was 20km away, of which had abundant trees and bird populations. Furthermore, to evaluate diversity and probable similarity in faecal source of contamination we included two sites (JHB S2 and JHB S1) located in Thembisa, Johannesburg, about 54 km away from Pretoria sites.

Generally, in antibiotic resistance BST, a database of antibiotic resistance patterns from known sources within an area is needed for isolates comparison. In this method sample level analysis or isolate-level analysis can be used. However, if it was assumed that a sample came from a single major source, as was the case in this study, then sample level analysis can be used (Wiggins et al., 2003). Nevertheless, it is very unlikely that a RHRW sample would contain only one source of contamination. Consequently, we also used isolate-level analysis which is suitable for analysis if samples are assumed to be contaminated by more than one source (Wiggins et al., 2003). We assumed that pigeon faecal isolated were representative of the major faecal sources including cows and chickens, since we had observed them feeding in close contact to faecal material from these animals. It should be noted that several studies on antibiotic resistance analysis (ARA) have cited it as a useful tool in assessing contamination sources with average rates of correct classification ranging from 62 to 84%, (Wiggins, 2003).

#### 6.4.1 Antibiotic resistance profiles of samples

In this study, antibiotic resistance was observed in *E. coli* isolates from all RHRW sites and bird faecal samples, similar to findings from previous studies (Saenz et al., 2003; Zahraei Salehi, 2005). The most frequently encountered form of resistance in all samples was to ampicillin (26.9%), gentamicin (26%), amoxicillin (25.2%), tetracycline (18.2%), and amikacin (17.8%) (Table 6.2). Most of these antibiotics have been widely used for therapeutic purposes against bacterial infections in humans and animals as well as growth promoters in agriculture and aquaculture (Kruse and Sorum, 1994; Khachatourians, 1998). Similar multidrug resistance phenotypes of *E. coli* isolated from water sources have been reported worldwide (Ozgumus et al., 2007; Torođlu and Dinçer, 2008).

The observation of relatively high antibiotic resistance percentages (Table 6.2) to similar antibiotics for *E. coli* from RHRW and birds, suggest birds to be the source of RHRW contamination. Similarly the observed correlation between pigeon faecal isolates and individual RHRW sites from both Pretoria and Johannesburg were all significant ( $r \geq 0.717$ ;  $p \leq 0.01$ ), except for isolates from PTA S3 ( $r=0.6$ ) in Pretoria. Isolates from the PTA S3 sites had no significant correlation to any of the other RHRW sites. This can be explained by site specific factors such as the presence of different species of birds or animals that may harbour *E. coli* with variant antibiotic resistance profiles. When we asked people residing at this site, about birds common in surrounding trees, no mention was given of pigeons. It is also important to note that except for this site we had seen pigeons in surrounding environments of all the other sites during sample collection. The PTA S3 site is located in a suburban area where garbage is efficiently collected and is not near garbage dump site, hence there is a limited supply of food. On the contrary the PTA S1 is situated on a farm near a cattle feed lot where thousands of pigeons come to feed. Although the PTA S2 site which is situated about 5 km from PTA S1, is on the second floor clear of trees, windblown bird dropping particles can explain the similarity. On the other hand, sites in Johannesburg are located in Thembisa, a high density township with a limited number of trees but abundant public refuse dump containers, and it was common to see dumped food and pigeons along the streets.

Isolates from birds had a relatively higher proportion of antibiotic resistance than those from RHRW tanks, with the most significant differences being noted in resistance to gentamicin (12.3%), tetracycline (9.8%), chloramphenicol (6.5%), and amikacin (4.2%) suggesting *E. coli* sources other than bird droppings only. This was supported by finding from Kruskal Wallis non-parametric tests of antibiotic resistance prevalence between isolates from RHRW and pigeon faeces, and between different RHRW sites, where significant differences ( $p \leq 0.05$ ) in 8 antibiotics were noted between all RHRW sites and pigeons and only one antibiotic between RHRW sites only. Hence it can be suggested that there may be other sources of faecal contamination at play, which harbour *E. coli* with less AR prevalence. Probable sources of these *E. coli* are natural environmental habitats, other species of birds, small animals such as rats and dust particles with *E. coli* populations that are not as exposed to antibiotic as are other domestic animals.

Given the erratic nature of rainfall in Gauteng and the amount of sun received, dust particles should be a significant factor of RHRW contamination. Previous research on the influence of prevailing air quality in the environment surrounding a rainwater harvesting site, have suggested air quality to play an important role in influencing the microbial quality of rainwater (Evans et al., 2006; Kaushik et al., 2012). The link between air quality and the microbial quality of rainwater under different weather conditions have been suggested as one of the reasons for discrepancies in reported literature concerning rainwater portability. For instance, research by Kaushik et al., (2012) on rainwater contamination factors that are independent of external sources of contamination such as those encountered with roof-harvesting methods found *E. coli* to be one of the microorganisms present in fresh rainwater and contributing to its microbial diversity. Furthermore, a study by Evans et al., (2006) on the influence of weather conditions on roof water contamination, found that wind speed and direction had a strong influence on the microbial quality of RHRW.

Similar to our findings, previous research have shown *E. coli* strains isolated from pigeon faeces to be antibiotic resistant and may reflect the abusive use of such substances in our society (Silva et al., 2009). Pigeons possess a less developed cecum, which is the part of the gastrointestinal tract that harbours the most abundant and diversified microbiota. Hence, it is possible that these microorganisms may not be true inhabitants of pigeon faecal flora.

Consequently, it is less likely that these birds could have a permanent gut microbiota with intestinal bacterial species commonly associated with humans (Baele et al., 2002). However, the feeding habits of urban pigeons and doves may include garbage from nearby trash cans, which exposes them to contamination with residual antimicrobials or chemicals and medically important bacteria (Haag-Wackernagel and Moch, 2004; Rosengren et al., 2009).

Despite the lack of reports on antimicrobial susceptibility patterns of *E. coli* from RHRW and their comparison to the likely sources of faecal origin, a number of researches have shown increasing resistance to tetracycline, ampicillin and gentamicin, with resistance at lower levels to cotrimoxazole, nalidixic acid chloramphenicol enrofloxacin cefoxitin and ciprofloxacin, similar to our findings (Kahlmeter, 2003; Cardonha et al., 2005; Tanaka et al., 2005).

#### **6.4.2 Multiple antibiotic resistance patterns**

High prevalence of MAR were noted on combinations of four antibiotics in isolates from JHB (18.8%) and birds 15.2% suggesting the JHB sites to be highly impacted by pigeon droppings. However, the low prevalence (5.1%) observed for PTA sites on the same number of MAR, suggests that birds may not be the only source of contamination for these sites. The most abundant resistance phenotype to four antibiotics (Ak-Gm-Cip-T) was dominated by isolates from birds (6.8%) with Pretoria and Johannesburg isolates having low proportions of 1.3 and 3.1% respectively.

It is important to note that most of the MAR observed included resistance to an aminoglycoside (Gentamicin (Gm) or amikacin (Ak) or both). Gentamicin, was originally approved for use in the USA in 1963 (Walsh, 2003) and is widely used in the poultry industry. In our study, 64.4% (47/73) of gentamicin-resistant *E. coli* isolates were multidrug resistant ( $\geq 3$  classes of drugs). Forty three percent (32/73) were resistant to  $\geq 4$  antibiotics, including 54.8% (40/73) to amikacin, 31.5% (23/73) to ampicillin, and 20.5% (15/73) to tetracycline and 32.9% 24/73 to ciprofloxacin.



Relatively high prevalence of resistance to gentamicin by both bird droppings and RHRW isolates can be explained by findings from previous research which suggested that selective pressure by heavy antibiotic usage may not be responsible for all common aminoglycoside resistance (Gardner, et al., 1969). It has also been shown that *E. coli* with aminoglycoside can multiply and persist in the gastrointestinal tract of poultry in the absence of selective pressure by antibiotics (Guilot et al., 1977). Furthermore, several genes that confer multiple-aminoglycoside resistance in *E. coli* have been reported (Davis et al., 2010). Hence, the observed prevalence in aminoglycoside resistance should not be surprising.

Other antibiotics of importance to which resistance was observed include tetracycline and chloramphenicol to a lesser extent. Tetracycline was approved in the USA in 1948 (McEwen and Fedorka-Cray, 2002; Walsh, 2003), and has widely been used in therapy and to promote feed efficiency in animal production systems. Persistence of tetracycline resistance in animal coliforms was first reported a decade after it was no longer used in animal feed or for treatment (Langlois et al., 1983). We commonly found co-resistance for tetracycline with amoxicillin, gentamicin, ciprofloxacin, cefoxitin, ampicillin, and chloramphenicol, as in other studies (Schroeder et al., 2002; Roberts, 2005; Kozak et al., 2009).

Another antibiotic, chloramphenicol, was approved in 1947 and was strictly used for human clinical purposes. However, florfenicol, a closely related drug, was approved for treatment of respiratory diseases in cattle (USDA, 2010). Consequently persistence of chloramphenicol resistance in *E. coli* has been observed (Sayah et al., 2005; Tadesse et al., 2012; Bischoff et al., 2005). The *flo* gene which mediates florfenicol resistance, confers nonenzymatic cross-resistance to chloramphenicol (Schwarz et al., 2004) and might select for nascent resistance in recent strains (Tadesse et al., 2012). It is important to note that only a small number of known chloramphenicol resistance genes mediate resistance to florfenicol (Schwarz et al., 2004). For instance, chloramphenicol-resistant strains in which resistance is exclusively based on activity of chloramphenicol acetyltransferases do not show resistance to florfenicol (Tadesse et al., 2012).

### 6.4.3 Cluster analysis of antibiotic resistance profiles

We further performed cluster analysis using data on inhibition zones diameters irrespective of whether the isolates were resistant or not, for all antibiotics used and isolates tested. Cluster analysis yielded a dendrogram (Figure 6.3), that grouped the 239 isolates into five (A, B, C D and E) major clusters at  $(Dlink/Dmax) \times 100 < 65$ , which subdivided into seventeen minor clusters (A1-E7) at  $(Dlink/Dmax) \times 100 < 10$  (Figure 6.3). The major findings from cluster analysis were similar to what we had previously noted with the analysis of AR prevalence of *E. coli* isolates to individual antibiotics between the different sites, areas and sources as well as by the observed multiple antibiotic resistance profiles. Isolates that were relatively unique to individual RHRW were observed for all sites and fifteen of the seventeen clusters had pigeon faecal isolates, suggesting pigeons to be the most likely source of contamination. A significant number of isolates from birds clustered on their own without any RHRW isolates present and vice versa. This observation suggests birds to harbour a wide diversity of *E. coli* than could be accounted for, from RHRW isolates. It is therefore imperative that the total potential impact of using RHRW be evaluated from the sources of potential contamination (i. e. pigeon faeces, etc.) so as to be in a position to evaluate the total risks and probable mitigation.

We also noted significant presence of isolates from both pigeon and RHRW isolates in particular clusters further confirming the notion that birds were the most likely source of contamination. From our observation during sampling, we saw pigeons feeding on garbage from human dwellings, chicken waste from a fowl run and cattle feed in a cattle feed lot. It is possible that *E. coli* population observed in the pigeons are a representation of all the sources where they feed, since it is most likely that they will consume faecal matter of other animals from whose waste and feed they depend on. However, this can only be concluded after an extensive study of the *E. coli* from all the probable sources. The observation of isolates that were unique to individual sites within an area or to the two different areas, suggest site specific and/or area specific factors of contamination. Since pigeons are the most likely source, we collected pigeon faecal sample from a farm in Pretoria where masses of birds congregate to feed on animal feed and waste. However, it is doubtful that faecal samples we collected could be representative of pigeons in Johannesburg and only an investigation into the diversity of *E. coli* from pigeons in the two cities can give the true diversity.

Nevertheless, we were able to find a significant number of isolates with common profiles from the two areas. Whether this was a result of convergent antibiotic resistance evolution or bird migration, only research can confirm.

## 6.5 Conclusion and recommendations for further studies

Our study has limitations, which included the inadequate number of isolates used to represent some of the sites. In general, individual culture-based diversity characterisation methods require a relatively large database to give a good community representation. These limitations result in an uneven distribution of isolates per site, area or source. Although the isolate selection can be considered to be random, the sample sizes were not the same for all the sites. Therefore, we selected the non-parametric tests of Kruskal–Wallis one-way analysis of variance for analyses of resistance as a function of the different sites, areas and sources, because they are suitable for non-normally distributed data and data with small number of observations.

Despite these limitations, this analysis provides foundational information on the antibiotic resistance diversity of *E. coli* present in RHRW. This lays the groundwork for understanding the multiple probable sources of contamination, and chances for remediation. These data show that urban pigeons, the most likely source of RHRW contamination are reservoirs of multiple antibiotic resistant bacteria. This is important, as RHRW is presumed safe and is generally consumed without prior treatment. This is of serious concern given that approximately 20% of the South African population is HIV positive (Shisana et al., 2009). Since the pathological implications of infection by *E. coli* harbouring pathogenic elements and multidrug resistance, especially on individuals with suppressed immune systems are devastating.

Currently, there is ongoing work to evaluate temporal and special variation in RHRW quality for different rainwater harvesting systems in various rural, urban and informal settlements in South Africa and its suitability for domestic use (WRC, 2012). This work will provide more definitive data on critical factors for sustainable RWH practices suitable for the different communities in South Africa. However, more work still need to be done to evaluate the

presence of antibiotic resistance genetic determinants and the presence of various pathogenic elements in the isolates. Further, work should also include not only birds, but all probable sources of contamination including weather factors such as wind and other sources of faecal pollution including domestic animals. Other microbes of importance such as *Salmonella*, *Campylobacter*, *Listeria*, enterococci and *Yersinia* should be included, together with microbial diversity characterisation techniques such as repetitive palindromic sequence (rep) PCR, single nucleotide polymorphism typing, phylogenetic and clade grouping. This will give more comprehensive data from which a relatively accurate estimation of the risks can be made, probable sources of contamination be determined and mitigation responses be appropriated with a better degree of accuracy and certainty.

## 6.6 References

- Abel, D. J. and W. T. Williams. 1985. A re-examination of four classificatory fusion strategies. *The Computer Journal* 28: 439-443.
- Ahmed, W., A. Vieritz, A. Goonetilleke and T. Gardner. 2010. Health risk from the use of roof harvested rainwater in Southeast Queensland, Australia, as potable or non-potable water, determined using quantitative microbial risk assessment. *Applied and Environmental Microbiology* 76: 7382–7391.
- Ahmed, W., F. Huygens, A. Goonetilleke, and T. Gardner. 2008. Real-time PCR detection of pathogenic microorganisms in roof harvested rainwater in Southeast Queensland, Australia. *Applied and Environmental Microbiology* 74: 5490–5496.
- Ahmed, W., L. Hodgers, N. Masters, J. P. S. Sidhu, M. Katouli and S. Toze. 2011b. occurrence of intestinal and extraintestinal virulence genes in *Escherichia coli* isolates from rainwater tanks in Southeast Queensland, *Applied and Environmental Microbiology* 77: 7394-7400.
- Ahmed, W., T. Gardner and S. Toze. 2011a. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Alekshun, M. N. and S. B. Levy. 1999. The *mar* regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends in Microbiology* 7: 410-413.

- Amin, M. T. and M. Y. Han. 2009. Roof harvested rainwater for potable purposes: application of solar collector disinfection (SOCO-DIS). *Water Research* 43: 5225–5235.
- Antai, S. P. 1987. Incidence of *Staphylococcus aureus*, coliforms and antibiotic resistant strains of *Escherichia coli* in rural water supplies in Port Harcourt. *The Journal of Applied Bacteriology* 62: 371–375.
- Baele, M., L. A. Devriese, P. Butaye and F. Haesebrouck. 2002. Composition of enterococcal and streptococcal flora from pigeon intestines. *Journal of Applied Microbiology* 92: 348–351.
- Barton, M. D. 1998. Does the use of antibiotics in animals affect human health? *Australian Veterinary Journal* 76:177-180.
- Bauer, A. W., W. M. Kirby, J. C. Sherris and M. Turck. 1966. Anti-biotic susceptibility testing by a standardised single disk method. *American Journal of Clinical Pathology* 45: 493–496.
- Bej, A. K., J. L. DiCesare, L. Haff and R. M. Atlas. 1991. Detection of *Escherichia coli* and *Shigella* spp. in water by using the polymerase chain reaction and gene probes for uid. *Applied and Environmental Microbiology* 57: 1013–1017.
- Bischoff, K. M., D. G. White, M. E. Hume, T. L. Poole and D. J. Nisbet. 2005. The chloramphenicol resistance gene *cmlA* is disseminated on transfer-able plasmids that confer multiple-drug resistance in swine *Escherichia coli*. *FEMS Microbiology Letters*. 243: 285–291.
- Bittar, F., Z. Ouchenane, F. Smati, D. Raoult, and J. M. Rolain. 2009. MALDI-TOF-MS for rapid detection of staphylococcal Panton-Valentine leukocidin. *International Journal of Antimicrobial Agents* 34: 467–470.
- Blanco, G., J. A. Lemus and J. Grande. 2009. Microbial pollution in wildlife: linking agricultural manuring and bacterial antibiotic resistance in red-billed choughs. *Environmental Research*. 109: 405-412.
- Cardonha, A. M., R. H. Vieira, G. Peirano and P. Rodrigues. 2005. Resistance to antibiotics and heavy metals from *Escherichia coli* isolated from sea water and pluvial galleries. *Journal Acta Cirurgica Brasileira* 20: 253–256.

- Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw. 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Applied and Environmental Microbiology* 67: 1503-1507
- Casewell, M., C. Friis, E. Marco, P. McMullin and I. Phillips. 2003. The European ban on growth-promoting antibiotics and emerging consequences for human and animal health. *Journal of Antimicrobial Chemotherapy*. 52: 159-161.
- Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of *Escherichia coli* phylogenetic group. *Applied and Environmental Microbiology*. 66: 4555-4558
- Clinical and Laboratory Standards Institute (CLSI). 2008. Performance standards for antimicrobial susceptibility testing, 8th informational supplement. Wayne, PA.
- Cole, D., D. J. V. Drum, D. E. Stallknecht, D. G. White, M. D. Lee, S. Ayers, M. Sobsey, and J. J. Maurer. 2005. Free-living Canada geese and antimicrobial resistance. *Emerging Infectious Diseases Journal* 11:935-938.
- Costa, D., P. Poeta, Y. Saenz, L. Vinue, B. Rojo-Bezares, A. Jouini, M. Zarazaga, J. Rodriguez, and C. Torres. 2006. Detection of *Escherichia coli* harbouring extended-spectrum  $\beta$ -lactamases of the CTX-M, TEM and SHV classes in faecal samples of wild animals in Portugal. *Journal of Antimicrobial Chemotherapy* 58: 1311-1312.
- Davis, M. A., K. N. Baker, L. H. Orfe, D. H. Shah, T. E. Besser and D. R. Call. 2010. Discovery of a Gene Conferring Multiple-Aminoglycoside Resistance in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy* 54: 2666-2669.
- Dobbin, G., H. Hariharan, P. Y. Daoust, S. Hariharan, S. Heaney, M. Coles, L. Price, and C. A. Muckle. 2005. Bacterial flora of free-living double-crested cormorant (*Phalacrocorax auritus*) chicks on Prince Edward Island, Canada, with reference to enteritic bacteria and antibiotic resistance. *Comparative Immunology, Microbiology and Infectious Diseases* 28: 71-82.
- Dolejska, M., B. Biersova, L. Kohoutova, I. Literak, and A. Cizek. 2009. Antibiotic resistant *Salmonella* and *Escherichia coli* isolates with integrons and extended-spectrum beta-lactamases in surface water and sympatric black-headed gulls. *Journal of Applied Microbiology* 106:1941-1950.

- Dolejska, M., D. Senk, A. Cizek, J. Rybarikova, O. Sychra, and I. Literak. 2008. Antimicrobial resistant *Escherichia coli* isolates in cattle and house sparrows on two Czech dairy farms. *Research in Veterinary Science* 85:491-494.
- Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Applied Environmental Microbiology* 66: 2572-2577.
- Donovan, E. P., D. F. Staskal, K. M. Unice, J. D. Roberts, L. C. Haws, B. L. Finley and M. A. Harris. 2008. Risk of gastrointestinal disease associated with exposure to pathogens in the sediments of the Lower Passaic River. *Applied and Environmental Microbiology* 74: 1004–1018.
- Evans, C. A., P. J. Coombes and R. H. Dunstan. 2006. Wind, rain and bacteria: the effect of weather on the microbial composition of roof harvested rainwater. *Water Research*. 40: 37–44.
- Fallacara, D. M., C. M. Monahan, T. Y. Morishita, and R. W. Wack. 2001. Faecal shedding and antimicrobial susceptibility of selected bacterial pathogens and a survey of intestinal parasites in free-living waterfowl. *Avian Diseases*. 45: 128-135.
- Gardner, P., D. H. Smith, H. Beer and R. C. Moellering, Jr. 1969. Recovery of resistance (R) factors from a drug-free community. *Lancet*. 2: 774-776.
- Guillot, J. F., E. Chaslus-Dancla and J. P. Lafont. 1977. Spontaneous implantation of antibiotic resistant Enterobacteriaceae in the digestive tract of chickens in the absence of selective pressure. *Antimicrobial Agents and Chemotherapy*. 12: 697-702.
- Haag-Wackernagel, D., and H. Moch. 2004. Health hazards posed by feral pigeons. *Journal of Infection*. 48: 307–313.
- Hachler, H., S. P. Cohen, and S. B. Levy. 1991. *marA* a regulated locus which controls expression of chromosomal multiple antibiotic resistance in *Escherichia coli*. *Journal of Bacteriology*. 173: 5532-5538.
- Hager, M. C. 2001. Detecting bacteria in coastal waters. Part 1. *The Journal for Surface Water Management Professionals* 2: 16–25.

- Harwood, V. J., J. Butler, D. Parrish, and V. Wagner. 1999. Isolation of faecal coliform bacteria from the diamondback terrapin (*Malaclemys terrapin centrata*). *Applied and Environmental Microbiology*. 65: 3698–3704.
- Harwood, V., B. Wiggins, B. Hagedorn, R. Ellender, J. Gooch, J. Kern and B. Thompson. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *Journal of Water and Health* 1: 153-166.
- Jacques, L., L. Thiago, V. L. Silva, J. R. Nicoli, T. C. Nascimento, and C. G. Diniz. 2009. Diarrheagenic *Escherichia coli* strains recovered from urban pigeons (*Columba livia*) in Brazil and their antimicrobial susceptibility patterns. *Current Microbiology* 59: 302–308.
- Kahinda, M. J., A. E. Taigbenu and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth* 32: 1050–1057.
- Kahlmeter, G. 2003. An international survey of the antimicrobial susceptibility of pathogens from uncomplicated urinary tract infections: The ECO. SENS Project. *Journal of Antimicrobial Chemotherapy*. 51: 69–76.
- Kaushik, R., R. Balasubramanian, and A. A. De la Cruz. 2012. Influence of air quality on the composition of microbial pathogens in fresh rainwater. *Applied and Environmental Microbiology* 78: 2813–2818.
- Khachatourians, G. G. 1998. Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *Canadian Medical Association Journal* 159: 1129-1136.
- Kinge, W. C. N., N. C. Njie and T. D. Kawadza. 2010. Antibiotic resistance profiles of *Escherichia coli* isolated from different water sources in the Mmabatho locality, North-West Province, South Africa. *South African Journal of Science* 106: 44–49.
- Kobayashi, H., T. Pohjanvirta and S. Pelkonen. 2002. Prevalence and characteristics of intimin and Shiga toxin-producing *Escherichia coli* from gulls, pigeons and broilers in Finland. *Journal Veterinary Medical Science* 64: 1071–1073.
- Kozak, G. K., D. L. Pearl, J. Parkman, R. J. Reid-Smith, A. Deckert, A and P. Boerlin. 2009. Distribution of sulfonamide resistance genes in *Escherichia coli* and *Salmonella*



- isolates from swine and chickens at abattoirs in Ontario and Québec, Canada. *Applied and Environmental Microbiology*. 75: 5999– 6001.
- Kruse, H. and Sorum, H. 1994. Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Applied and Environmental Microbiology* 60: 4015-4021.
- Langlois, B. E., G. L. Cromwell, T. S. Stahly, K. A. Dawson, V. W. Hays. 1983. Antibiotic resistance of faecal coliforms after long-term withdrawal of therapeutic and sub-therapeutic antibiotic use in a swine herd. *Applied and Environmental Microbiology*. 46: 1433–1434.
- Lee, J., J. Yang, M. Han and J. Choi. 2010. Comparison of the microbiological and chemical characterisation of harvested rainwater and reservoir water as alternative water resources. *Science of the Total Environment*. 408: 896–905.
- Levy S. B. 1992. Active efflux mechanisms for antimicrobial resistance. *Antimicrobial Agents and Chemotherapy*. 36: 695–703.
- Lewis, J. S., M. Herrera, B. Wickes, and J. E. Petterson. 2007. First report of the emergence of CTX-M-type extended-spectrum  $\beta$ -lactamases (ESBLs) as the predominant ESBL isolated in a U. S. health care system. *Antimicrobial Agents and Chemotherapy*. 51: 4015-4021.
- Literak, I., R. Vanko, M. Dolejska, A. Cizek, and R. Karpiskova. 2007. Antibiotic resistant *Escherichia coli* and *Salmonella* in Russian rooks (*Corvus frugilegus*) wintering in the Czech Republic. *Letters in Applied Microbiology*. 45: 616-621.
- Livermore, D. M., M. Warner, L. M. C. Hall, V. I. Enne, S. J. Projan, P. M. Dunman, S. L. Wooster, and G. Harrison. 2001. Antibiotic resistance in bacteria from magpies (*Pica pica*) and rabbits (*Oryctolagus cuniculus*) from west Wales. *Environmental Microbiology*. 3: 658-661.
- Marques, S. M. T., R. M. de Cuadros, C. J. Silva, and M. Baldo. 2007. Parasites of pigeons (*Columba livia*) in urban areas of lages, Southern Brazil. *Parasitología latinoamericana*. 62: 183–187.
- Martin, A. R., P. J. Coombes, T. L. Harrison and R. H. Dunstan. 2010. Changes in abundance of heterotrophic and coliform bacteria resident in stored water bodies in relation to

- incoming bacterial loads following rain events. *Journal of Environmental Monitoring*. 12: 255–260.
- McEwen, S. A. and P. Fedorka-Cray. 2002. Antimicrobial use and resistance in animals. *Clinical Infectious Diseases* 34: 93–106.
- Meays, C. L., K. Broersma, R. Nordin and A. Mazumder. 2004. Source tracking faecal bacteria in water: A critical review of current methods. *Journal of Environmental Management* 73: 71–79.
- Mohapatra, B. R., K. Broersma and A. Mazumder. 2007. Comparison of five rep-PCR genomic fingerprinting methods for differentiation of faecal *Escherichia coli* from humans, poultry and wild birds. *FEMS Microbiology Letters* 277: 98–106.
- Nawaz, M., M. Y. Han, T. Kim, U. Manzoor and M. T. Amin. 2012. Silver disinfection of *Pseudomonas aeruginosa* and *E. coli* in rooftop harvested rainwater for potable purposes. *Science of the Total Environment*. 431: 20–25.
- Oliveira, A. J. F. C., and J. M. W. Pinhata. 2008. Antimicrobial resistance and species composition of enterococci isolated from waters and sands of marine recreational beaches in Southeastern Brazil. *Water Research*. 42: 2242–2250.
- Ozgunus, O. B., E. Celik-Sevim, S. Alpay-Karaoglu, C. Sandalli and A. Sevim. 2007. Molecular characterization of antibiotic resistant *Escherichia coli* strains isolated from tap and spring waters in a Coastal Region in Turkey. *Journal of Microbiology* 45: 379–387.
- Pinfold, J. V., N. J. Horan, W. Wiroganagud, and D. Mara. 1993. The bacteriological quality of rainjar water in rural northeast Thailand. *Water Research*. 41: 2039–2047.
- Pinto, L., P. Poeta, H. Radhouani, C. Coelho, C. Carvalho, J. Rodrigues, C. Torres, R. Vitorino, P. Domingues and G. Igrejas. 2011. Proteomic study in an *Escherichia coli* strain from seagulls of the Berlengas Natural Reserve of Portugal. *Genetics* 1: 36–41.
- Power, M. L., J. Littlefield-Wyer, D. M. Gordon, D. A. Veal, and M. B. Slade. 2005. Phenotypic and genotypic characterisation of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environmental Microbiology* 7: 631–640.
- Prescott J. F., J. D. Baggot, and R. D. Walker (ed.) 2000. *Antimicrobial therapy in veterinary epidemiology*, 3rd ed. Iowa State University Press, Ames.

- Radimersky, T., P. Frolkova, D. Janoszowska, M. Dolejska, P. Svec, E. Roubalova, P. Cikova, A. Cizek, and I. Literak. 2010. Antibiotic resistance in faecal bacteria (*Escherichia coli*, enterococci) in feral pigeons. *Journal of Applied Microbiology* 109: 1687-1695.
- Richmond, M. H. 1972. Some environmental consequences of the use of antibiotics: or “what goes up must come down. *Journal of Applied Bacteriology*. 35: 155-176.
- Roberts, M. C. 2005. Update on acquired tetracycline resistance genes. *FEMS Microbiology Letters*. 245: 195–203.
- Romanus, I. I., A. T. Eze, O. A. Egwu, A. N. Felicitas and N. A. Chidiebube. 2011. Comparison of matrix-assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF-MS) with conventional culture and biochemical method of bacteria identification to species level. *Journal of Medical Laboratory and Diagnosis* 2: 1–4.
- Romesburg, H. C. 1990. *Cluster analysis for researchers*. Robert E. Kreiger Publishing Company, Inc. Malabar, Florida.
- Rose, J. M., R. J. Gast, A. Bogomolni, J. C. Ellis, B. J. Lentell, K. Touhey, and M. Moore. 2009. Occurrence and patterns of antibiotic resistance in vertebrates off the Northeastern United States coast. *FEMS Microbiology Ecology* 67: 421-431.
- Rosengren, L. B., C. L. Waldner and R. J. Reid-Smith. 2009. Associations between antimicrobial resistance phenotypes, antimicrobial resistance genes and virulence genes of faecal *Escherichia coli* from healthy grow-finish pigs. *Applied and Environmental Microbiology* 75: 1373– 1380.
- Rubens, C. E., W. F. McNeill, and W. E. Farrar, Jr. 1979. Evolution of multiple antibiotic resistance plasmids mediated by transposable plasmid deoxyribonucleic acid sequences. *Journal of Bacteriology* 140: 713-719.
- Rybarikova, J., M. Dolejska, D. Materna, I. Literak, and A. Cizek. 2010. Phenotypic and genotypic characteristics of antimicrobial resistant *Escherichia coli* isolated from symbovine flies, cattle and sympatric insectivorous house martins from a farm in the Czech Republic (2006-2007). *Research in Veterinary Science* 89:179-183.

- Saenz, Y., M. Zarazaga, L. Brinas, F. Ruiz-Larrea and C. Torres, 2003. Mutations in *gyrA* and *parC* genes in nalidixic acidresistant *Escherichia coli* strains from food products, humans and animals. *Journal of Antimicrobial Chemotherapy* 51: 1001-1005.
- Sayah, R. S., J. B. Kaneene, Y. Johnson and R. Miller. 2005. Patterns of antimicrobial resistance observed in *Escherichia coli* isolates obtained from domestic- and wild-animal faecal samples, human septage, and surface water. *Applied and Environmental Microbiology* 71: 1394–1404.
- Sazakli, E., A. Alexopolous, and M. Leotsinidis. 2007. Rainwater harvesting quality assessment and utilisation in Kelafonia Island, Greece. *Water Research* 41: 2039-2047.
- Schroeder, C. M., C. Chao, C. DebRoy, J. Torcolini, S. Zhao and D. G. White. 2002. Antimicrobial resistance of *Escherichia coli* O157 isolated from humans, cattle, swine, and food. *Applied and Environmental Microbiology* 68: 576–581.
- Schwarz, S., C. Kehrenberg, B. Doublet and A. Cloeckaert. 2004. Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiology Review* 28: 519–542.
- Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo and J. Lukasik. 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. *Applied and Environmental Microbiology*. 69: 1089–1092.
- Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik. 2002. Microbial source tracking: current methodology and future directions. *Applied and Environmental Microbiology* 68: 5796-5803.
- Shisana, O., L. Simbay, T. Rehle, N. P. Zungu, K. Zuma, N. Ngogo, S. Jooste, V. P van Wyk, W. Parker, S. Pezi, A. O. Davids, T. H. Dinh. 2010. South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2008: The health of our children. HSRC Press, Cape Town
- Silva, V. L., J. R. Nicoli, T. C., Nascimento, and C. G. Diniz. 2009. Diarrheagenic *Escherichia coli* strains recovered from urban pigeons (*Columba livia*) in Brazil and their antimicrobial susceptibility patterns. *Current Microbiology* 59: 302-308.

- Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner. 2002. Microbial source tracking: state of the science. *Environmental Science and Technology* 36: 5279-5288
- Singh K. P., A. Malik, D. Mohan and S. Sinha. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study. *Water Research* 38: 3980 -3992.
- Spinks J., S. Phillips, P. Robinson and P. Van Buynder. 2006. Bushfires and tank rainwater quality: A cause for concern. *Journal of Water and Health* 4: 21–28.
- Spratt, B. G. 1994. Resistance to antibiotics mediated by target alterations. *Science* 264: 388-393.
- Stevens, J. 1996. Discriminant analysis. In *Applied multivariate statistics for the social sciences*. pp 232-267. Lawrence Erlbaum Associates inc, Hillsdale, N. J.
- Stockholm Environment Institute (SEI). 2009. Rainwater harvesting: a lifeline for human well-being. United Nations Environmental Program. Job No: DEP/1162/NA UNEP.
- Sung, M., C. C. Kan, M. W. Wan, C. R. Yang, J. C. Wang, K. C. Yu and S. Z. Lee. 2010. Rainwater harvesting in schools in Taiwan: system characteristics and water quality. *Water Science and Technology* 61: 1767–1778.
- Tadesse, D. A., S. Zhao, E. Tong, S. Ayers, A. Singh, M. J. Bartholomew and P. F. McDermott. 2012. Antimicrobial drug resistance in *Escherichia coli* from humans and food animals. *Emerging Infectious Diseases* 18: 741–749.
- Tanaka, C., T. Miyazawa, M. Watari and N. Ishiguro. 2005. Bacteriological survey of faeces from feral pigeons in Japan. *Journal Veterinary Medical Science*. 67: 951–953.
- Torođlu, S. and S. Dincer. 2008. Elimination and conjugal transfer of their resistance in isolated gram negative bacteria from Aksu River (Kahramanmara<sup>o</sup>- Turkey). *Advances in Environmental Biology* 2: 124-131.
- U. S. Environmental Protection Agency. 2002. Health risks from microbial growth and biofilms in drinking water distribution systems. U. S. Environmental Protection Agency Office of Ground Water and Drinking Water Standards and Risk Management, Pennsylvania, Washington DC

- Van den Bogaard, A. E., N. London, C. Driessen and E. E. Stobberingh. 2001. Antibiotic resistance of faecal *Escherichia coli* in poultry, poultry farmers and poultry slaughterers. *Journal of Antimicrobial Chemotherapy*. 47: 763–771.
- Walsh, C. T. 2003. *Antibiotics: actions, origins, resistance*. Washington DC: American Society for Microbiology Press.
- Wani, S. A., L. Samanta, M. A. Bhat and Y. Nishikawa. 2004. Investigation of Shiga toxin-producing *Escherichia coli* in avian species in India. *Letters in Applied Microbiology* 39: 389–394.
- Ward, J. H. 1963. Hierarchical grouping to optimise an objective function. *Journal of the American Statistics Association*. 58: 236-244
- Whitman, R. L., and M. B. Nevers. 2003. Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. *Applied and Environmental Microbiology* 69: 5555–5562.
- Wiggins, B. A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner. 2003. Use of antibiotic resistance analysis for representativeness testing of multi- watershed libraries. *Applied and Environmental Microbiology* 69: 3399–3405.
- Witte, W. 1998. Medical consequences of antibiotic use in agriculture. *Science* 279: 996-997.
- Wunderlin, D. A., M. P. Diaz, M. V. Ame, S. F. Pesce, A. C. Hued and M. A. Bistoni. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). *Water Research* 35: 2881-2894.
- Zahraei Salehi, T., M. R. Mahzounieh and A. Saeedzadeh. 2005. The isolation of antibiotic-resistant salmonella from intestine and liver of poultry in Shiraz Province of Iran. *International Journal of Poultry Science* 4: 320-322.

## CHAPTER SEVEN

### PYROSEQUENCING ANALYSIS OF ROOF HARVESTED RAINWATER AND RIVER WATER USED FOR DOMESTIC PURPOSES IN LUTHENGELE VILLAGE IN THE EASTERN CAPE PROVINCE IN SOUTH AFRICA

#### Abstract

Roof harvested rainwater (RHRW) and alternative water sources used by rural communities in South Africa are prone to contamination by a variety of pathogens through faecal matter. In order to better understand the microbiological quality of RHRW, pyrosequencing analysis was used to study a single site model. Pyrosequencing targeting the V1-V3 hypervariable of the 16S rDNA, was used to investigate the bacterial diversity in river and RHRW used for potable purposes by rural households in Luthengele village in the Eastern Cape Province of South Africa. The generated sequence libraries contained on average 1,265 reads (range 855–1,739 reads) of 16S rDNA sequences per sample. The phylum *Proteobacteria* dominated the data set (80.5% of all reads), while 4.19% 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* (50.35% of all reads), *Alphaproteobacteria* (16.19%), *Verrucomicrobiae* (6.59%), *Planctomycetacia* (5.69%) and *Sphingobacteria* (3%) dominated the data set in all the samples. Although the class *Verrucomicrobiae* constituted 6.59% of all sequences, 88.53% of the sequences were from the river sample where the class represented 43.74% 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

## 7.1 Introduction

In most South African rural areas and informal settlements municipal water is either inconsistently supplied or is not available at all. Hence, people rely on RHRW for domestic purposes (Kahinda et al., 2008, 2010). It has therefore become necessary that RHRW becomes an essential element of the current water management practices. However, the use of RHRW is limited, mainly by the uncertainty over the microbial quality (Ahmed et al., 2011). Although, there is a general public health perception that RHRW is safe to drink, the presence of potential pathogens such as *Escherichia coli*, *Aeromonas* spp., *Campylobacter* spp., *Salmonella* spp., *Legionella pneumophila*, *Giardia* spp., *Giardia lamblia*, *Cryptosporidium* spp., *Vibrio* spp. and enteric viruses, have been reported in this water source (Ahmed et al., 2011). Rainwater can potentially be contaminated from faecal matter that get deposited on the roof catchment surface and is washed into the storage tanks during rainfall events (Ahmed et al., 2010). While efforts are ongoing to minimise contamination in RHRW, successful mitigation is limited unless the variety and concentrations of contaminating pathogens are known (Dorfman et al., 2009). It is therefore important to develop tools that can effectively detect and characterise pathogens of concern in RHRW.

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., 2012). 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 limiting the extent to which one can fully understand the potential infective risks from RHRW use. Potential exposure routes to these pathogens maybe through ingestion, inhalation, and dermal contact, during consumption, washing, bathing, domestic food gardening and irrigation (Ahmed et al., 2010b).

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. 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., 2012). Inside the storage tank, microorganisms may die with time or survive and proliferate in biofilms (Rasid, 2009). Bacteria make up a large portion of the microorganisms that can be present in RHRW 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., 2012c).

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 with the environmental factors is critical in understanding factors influencing HRW quality. In particular sources of microbial pollution in the form of faecal material, dust and biofilm formation and proliferation in the tank are critical in understanding the observed microbial population in RHRW (Ahmed et al., 2010b).

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, microbial communities in water 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  $10^3$  sequences identifications per sample (Hayashi et al., 2001; DeSantis et al., 2007). This sampling depth is not good enough to identify less common pathogens, which are typically present at concentrations far less than 1% proportion of the environmental microbial population (Bibby et al., 2010). The use of culture-independent molecular methods to characterise microbial communities have gained popularity because more than 90% of the microbes present in the environment are difficult to culture or are unculturable (Schloss and Handelsman, 2005). Sequence analysis of the 16S ribosomal RNA (16S rRNA) gene forms

the basis of most molecular techniques used for microbial identification. The suitability of the 16S rRNA gene for microbial identification emanates from it being universal in bacteria (Tringe and Hugenholtz, 2008). The 16S rRNA gene is characterised by extremely conserved regions that are interspersed by nine (V1-V9) species dependant hypervariable regions. Although the standard for accurate bacteria classification requires sequencing of the whole 16S rRNA gene, a number of studies have reported accurate microbial taxonomic analysis using hypervariable regions of the 16S rRNA (Huse et al., 2008; Claesson et al., 2009; Wang and Qian, 2009).

Traditionally, microbial community analysis of the 16S rRNA would typically involve capillary electrophoresis of a few hundred cloned sequences using Sanger sequencing technology. In these analyses, only the dominant phylotypes within the target population would be characterised. 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). Initially, the pyrosequencing technique could only achieve sequence reads of 250bp or less (Sanapareddy et al., 2009). However, with technology improvement, the read length now averages up to 500bp. The longer read length allows for a more definitive phylogenetic based identification and classification of individual reads.

Application of high-throughput pyrosequencing technologies has dramatically reduced the cost and time of sequencing and enabled the generation of several thousand sequences per sample (Huse et al., 2008). Not only does the technology allow for the detection and identification of dominant bacterial phylotype profiles within a sample but the high sequence numbers produced allows for the detection of rare species including pathogens within bacterial communities. Furthermore, the large number of 16S rRNA genes analysed per sample allow for a more accurate representation of the relative abundances of the bacterial phylotypes present (Sogin et al., 2006; Huse et al., 2008). 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). However, no report could be found on RHRW microbial communities. The aim of this study was to assess microbial community structures 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 pyrosequencing analysis technique to detect potential pathogens in RHRW was evaluated.

## **7.2 Methodology**

### **7.2.1.1 Study site and sample collection**

A total of nine water samples were collected in 2L plastic containers in duplicates directly from RHRW tanks (seven) and two individual samples from river water (one) 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 practices on RHRW microbiological quality. The study site, Luthengele village in the Eastern Cape Province of South Africa is situated in a mountainous area and has steep terrain. 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.

### **7.2.2 DNA extraction and Pyrosequencing**

Bacteria in water samples (1L) were concentrated onto 0.45µm cellulose nitrate filters (Sartorius Stedim Biotech, Goettingen, Germany) by vacuum filtration. Filters were stored at -20°C for less than 2 weeks prior to DNA extraction and pyrosequencing analysis. DNA was extracted from filters using the Quick-gDNA Miniprep kit (ZymoResearch, USA) using the manufacturer's protocol. DNA samples were sent for tag encoded pyrosequencing at Inqaba Labs (Pretoria, South Africa) using the Roche GS FLX+ 454 pyrosequencer (GATC Biotech,

Germany). The 16S rRNA gene was amplified using universal bacterial primer set 27f (5'-GAGTTTGATCMTGGCTCAG-3') and 1492r (5'-AAGGAGGTGATCCANCCRCA-3').

### 7.2.3 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. The webserver SnoWMAAn 1.15 (<http://snowman.genome.tugraz.at>) (Stocker et al., 2011) was used with the following settings for the taxonomic-based analysis: analysis type: UCLUST pipeline; chimera filtering reference database: Chimerarefdfs\_Feb\_2012\_aligned; Mothur alignment reference database: fungalITS\_24-Sep-2013-current\_prokMSA\_aligned; Mothur chimera checking reference database: chimerarefdfs\_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).

### 7.2.4 Pathogen identification

Sorted and trimmed sequences that had been classified to belong to pathogenic genera at 80% confidence threshold were selected to identify pathogens. These genera classifications were examined for the presence of sequences of pathogenic species known to be present in water (Rusin et al., 1997; World Health Organisation, 2004). 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 methods 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) .

## **7.3 Results**

### **7.3.1 Sequencing depth**

The sequencing of the 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 unquing step and were used for alignment and chimera filtering, 3101 (22.12%) sequences were identified as chimeras and were excluded from downstream analysis, leaving 10956 sequences. On average, an individual sample harboured 1,265 reads (range 855 –1,739 reads).

The 10 956 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 level.

### **7.3.2 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 read length of at least 150bp per sample. At 20% sequence divergence, rarefaction curves for all samples showed saturation, indicating that the surveying effort covered almost the full extent of phylum level taxonomic diversity (Figure 7.1). 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 bacterial 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.

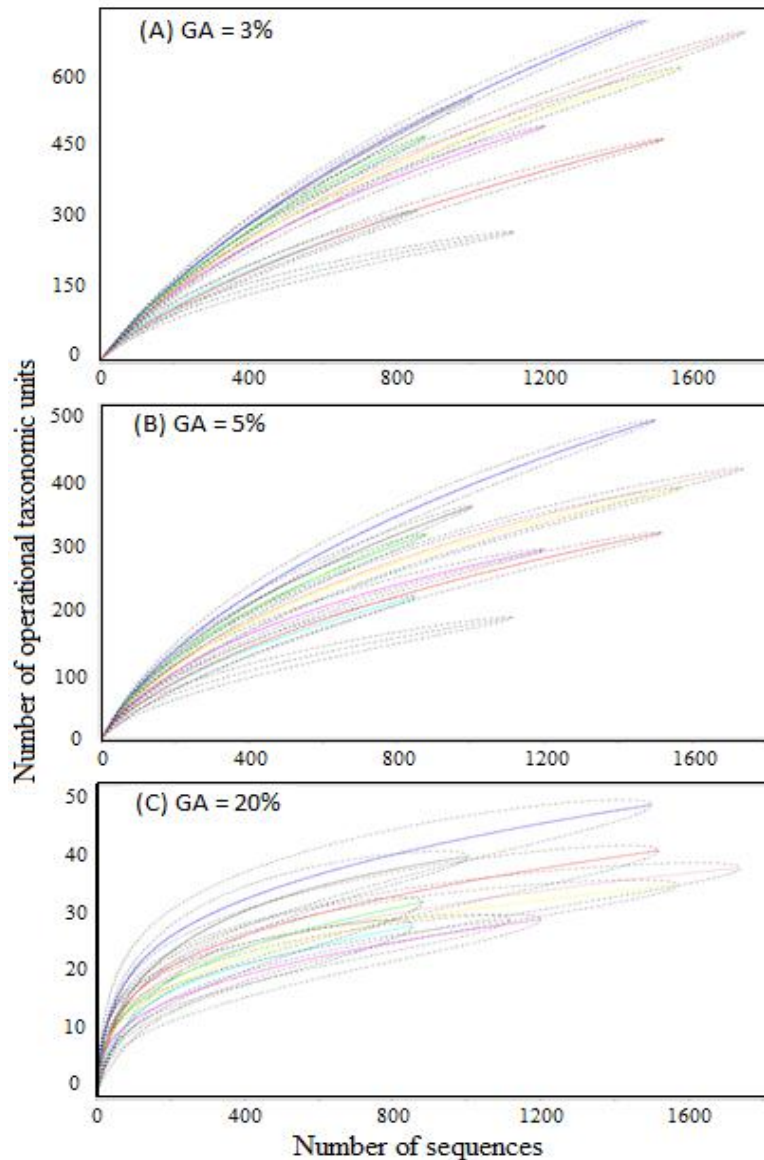


Figure 7. 1: Rarefaction curves for species richness and diversity at genetic distances of 3% (A) (species level), 5% (B) (genus level) and 20% (C) (phylum level).

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 7.1). 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 7. 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

### 7.3.3 Phylum level diversity

At phylum level 95.6% of sequences were classified into ten phyla of the domain bacteria (Figure 7.2). *Proteobacteria* dominated the data set (80.5% of all reads), while 4.19% of the reads could not be classified to any of the known phyla at a probability of 0.8 or higher (unclassified bacteria), 8.68% of reads were classified as *Planctomycetes*, 4.1% as *Bacteroidetes*, 0.73% as *Acidobacteria*, 0.73% as *Verrucomicrobia*, 0.53% as candidate division OP10, 0.46% as *Actinobacteria*, 0.07% as *Nitrospira*, 0.02% as *Firmicutes* and 0.01% as *Gemmatimonadetes*.

*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 *OP10* 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 (1.79%).

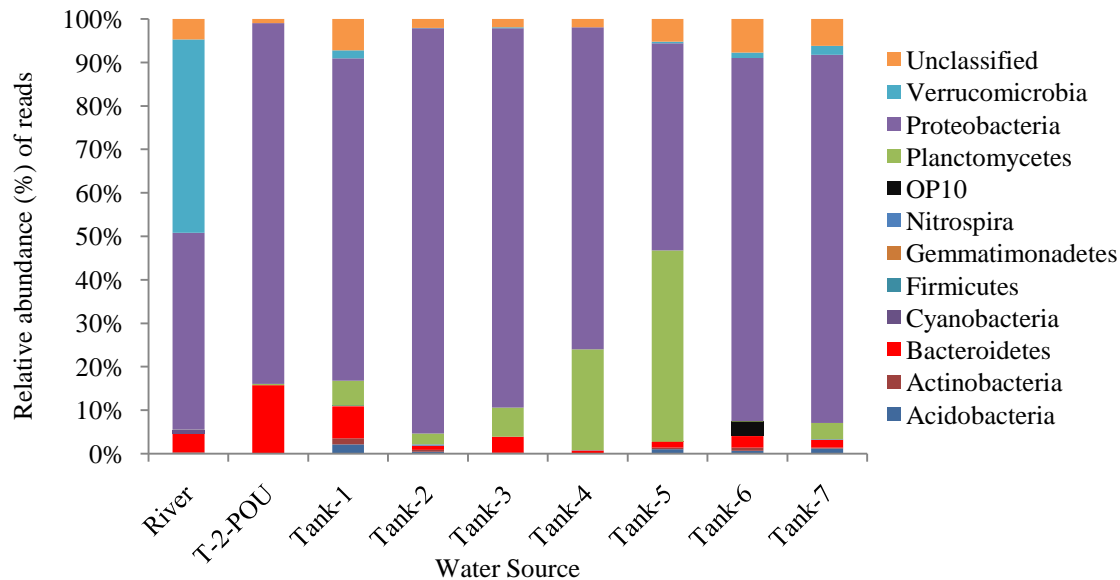


Figure 7. 2: Average relative abundance (%) of the reads classified into the predominant microbial phyla by individual samples.

### 7.3.4 Class level diversity

At class level, 85.6% of sequences were classified into sixteen classes (Figure 7.3). The classes; *Betaproteobacteria* (50.35% of all reads), *Alphaproteobacteria* (16.19%) *Verrucomicrobiae* (6.59%), *Planctomycetacia* (5.69%) and *Sphingobacteria* (3%) dominated the data set in all the samples. Of the sixteen classes, 10 had lower relative abundance ratio (<1%). These classes had either low relative abundances in most of the samples i. e. *Gammaproteobacteria* (0.82%), *Acidobacteria* (0.53%), *Actinobacteria* (0.38%) or were absent. Although the class *Verrucomicrobiae* constituted 6.59% of all sequences, 88.53% of the sequences were from the river sample where the class represented 43.74% of the observed sequences in the sample. The class *Flavobacteria* was detected in high abundance in the kitchen sample (6.45%), although it was also detected in river, tank 1 and tank 6 samples and it represented less than 1% of the observed sequences in these samples. 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 *OP10 genera incertae sedis* detected only in tank 3 (0.13%) and tank 6 (3.35%).

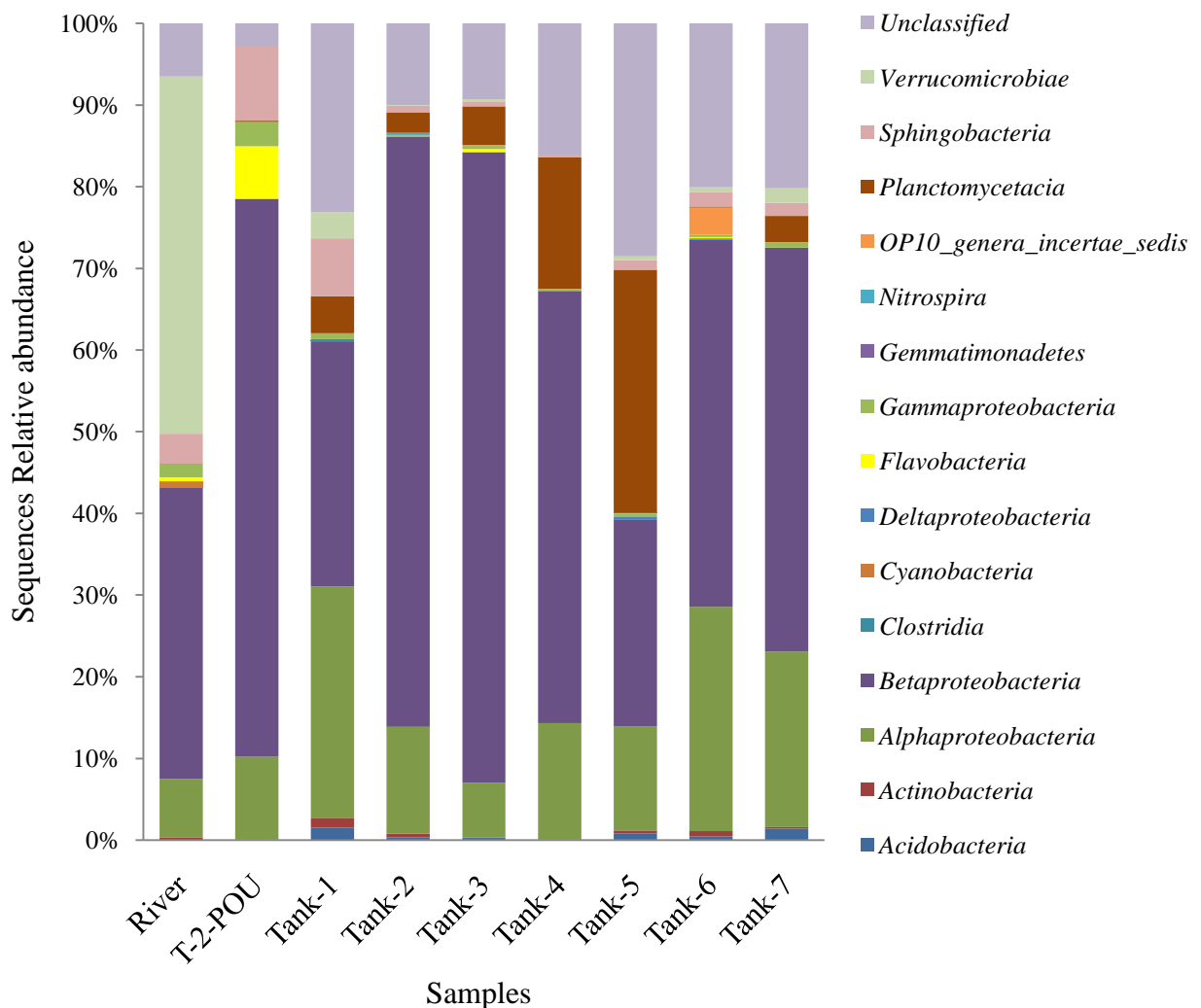


Figure 7. 3: Average relative abundance (%) of the reads classified into the predominant microbial classes of individual samples.

### 7.3.5 Diversity at order level

At order level 67.93% of sequences were classified into twenty-four orders. However, of the 67.93 classified sequences, 91.18% belonged to five orders; *Burkholderiales* (64.17%), *Verrucomicrobiales* (9.69%), *Planctomycetales* (8.38%), *Sphingobacteriales* (4.38%), *Sphingomonadales* (4.55%), *Rhizobiales* (1.64%) (Figure 7.4). Other orders that had abundance frequencies >1% were *Rhizobiales* (1.64%), *Rhodobacterales* (1.29%) and *Flavobacteriales* (1.14%). There were large variations in the relative abundances of a

number of orders between 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 (12.54%) samples, *Planctomycetales* in tank 4 (22.21%) and 5 (55.46%). Wide relative abundance variation of particular orders in individual samples were observed for the orders *Burkholderiales* (32.75-87.34%), *Verrucomicrobiales* (0.16-48.22%), *Planctomycetales* (0.07-55.46%). Orders that were not widely distributed among the samples included Chloroplast that was 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 (2.81%) samples.

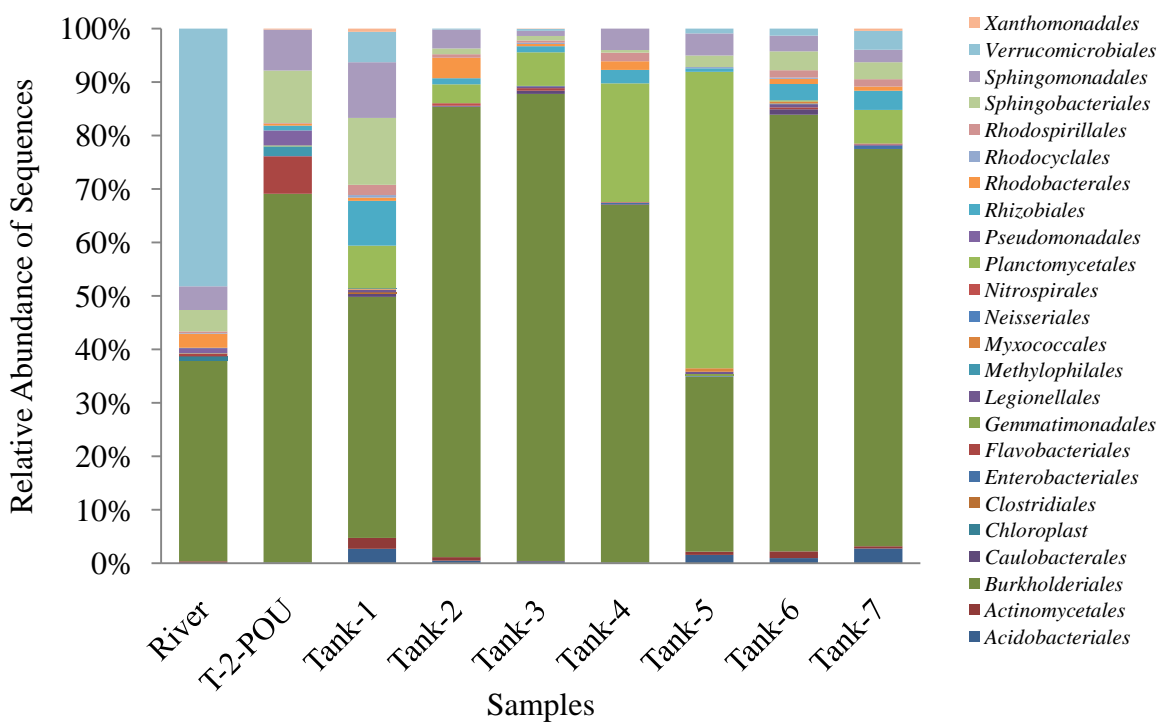


Figure 7. 4: Average relative abundance (%) of the reads classified into the predominant microbial classes of individual samples.

### 7.3.6 Diversity at family level

At family level 41.55% of sequences could be classified into thirty-seven families of which the most abundant were *Comamonadaceae* (33.1%), *Oxalobacteraceae* (16.57%), *Planctomycetaceae* (13.7%), *Sphingomonadaceae* (7.44%) (Figure 7.5). Other abundant families included *Genera Incertae sedis* 5, *Burkholderiaceae*, *Crenotrichaceae*,

*Flexibacteraceae*, *Rhodobacteraceae*, *Flavobacteriaceae*, *Subdivision 3*, *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* (14.61-60.09%), *Burkholderiaceae* (0.59-26.9%), *Planctomycetaceae* (0.41-63.98%) and *Crenotrichaceae* (0.43-13.22%). Families that were unique to particular samples included *Opitutaceae* (0.54%), *Bacillariophyta* (2%) and *Verrucomicrobiaceae* (9.61%) in river sample; *Moraxellaceae* in kitchen sample (2.91%); *Microbacteriaceae* in tank 1 (0.48%) and tank 5 (0.76%); *Pseudomonadaceae* in river (2.36%) and kitchen (0.34%).

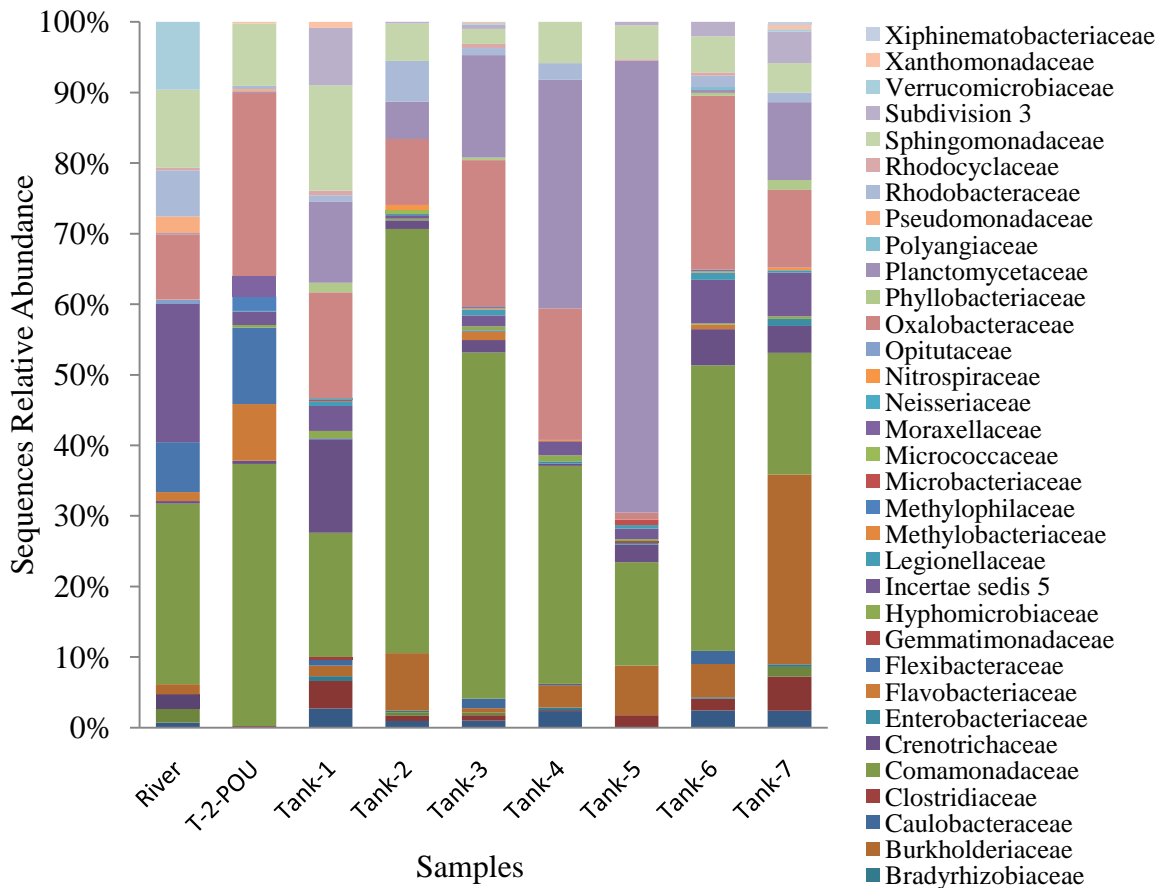


Figure 7. 5: Relative abundance (%) of sequences classified into respective bacteria families of the individual samples.

#### 7.4 Cluster analysis and genus level diversity

Cluster analysis was performed in order to group the samples into similarity groups and reduce the number of individual sample in the analysis. This would enable in-depth analysis of the observed variation among the samples and to detect trends that characterises the observed microbial community groups.

### 7.4.1 Cluster analysis

Cluster analysis of the nine samples at the OTU-level data produced the highest resolution for differentiating the three water sample groups bacterial communities. At a similarity distance (SD) of ~240 (SD240) the generated cluster diagram grouped the samples into river and RHRW sources. At SD ~165, the RHRW samples were grouped into kitchen and tank water. The tank water group comprised of seven tank water samples collected from seven different households whereas river and kitchen water were comprised of individual samples. The tank water sample group clustered between SD ~70 and SD ~120. The differences between kitchen and tank water samples in the cluster diagram of microbial communities based on observed OTUs suggests that the process of fetching and storing tank water prior to use, significantly alters the microbial population structure of the water. Where grouping of samples was necessary during the analysis, we grouped the samples into three groups (river, kitchen and tank) as we had observed from the cluster analysis dendrogram (Figure 7.6).

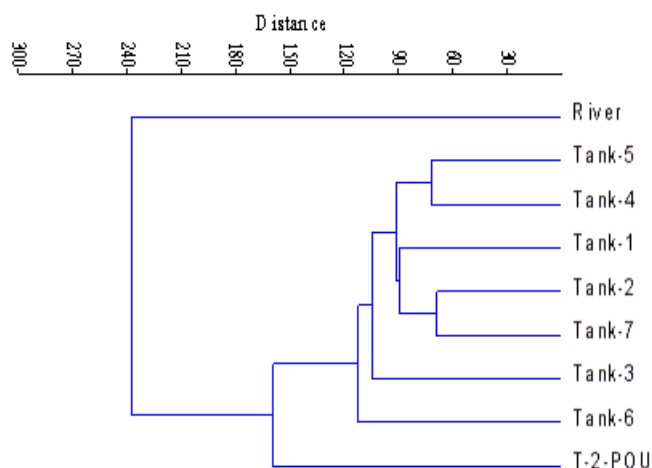


Figure 7. 6: Clustering of samples. Bray-Curtis similarity index was calculated using the abundance of OTUs and hierarchical clustering was calculated using the single linkage and Euclidean algorithms

### 7.4.2 Genus level diversity

The differences in microbial community composition were made obvious by analysing the genera shared among the three sample groups. The use of single river and kitchen water samples to compare against seven RHRW samples may lead to biased results due to relatively low representation of sequence reads from single sample represented groups. However, the detection of shared and unique species in river and kitchen water samples compared to RHRW samples is valid in showing the differences in microbial quality among the three groups. Uclust's Venn diagram analysis was used to identify the genera with shared OTUs among samples (3% cut-off) as shown in Figure 7.7. In the Venn diagram, seven tank water samples 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 7.8). 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 7.8 also indicates that tank water samples contained a greater variety of bacteria (50 genera) than the river water sample (38) genera) and 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) and followed by tank and kitchen water (5 genera).

A comparison of the relative abundances of the classified bacterial genera in the water samples confirmed differences between the samples (Figure 7.8). Most notably, *Curvibacter* was the most abundant genus on all sampling sites. In river water (Figure 7.8A), the most frequently present genus was *Arcinella* (12.6%) and *Curvibacter* (19.9%) among the other

commonly found genera *Acidovorax*, *Aquabacterium*, *Cellvibrio*, *Janthinobacterium*, *Novosphingobium*, *Prostheco bacter* and *Pseudorhodobacter*. The dominant genera in kitchen water (Figure 7.8B) were *Curvibacter* (23%), *Janthinoacterium* (19%), *Acidovorax* (13.2%)

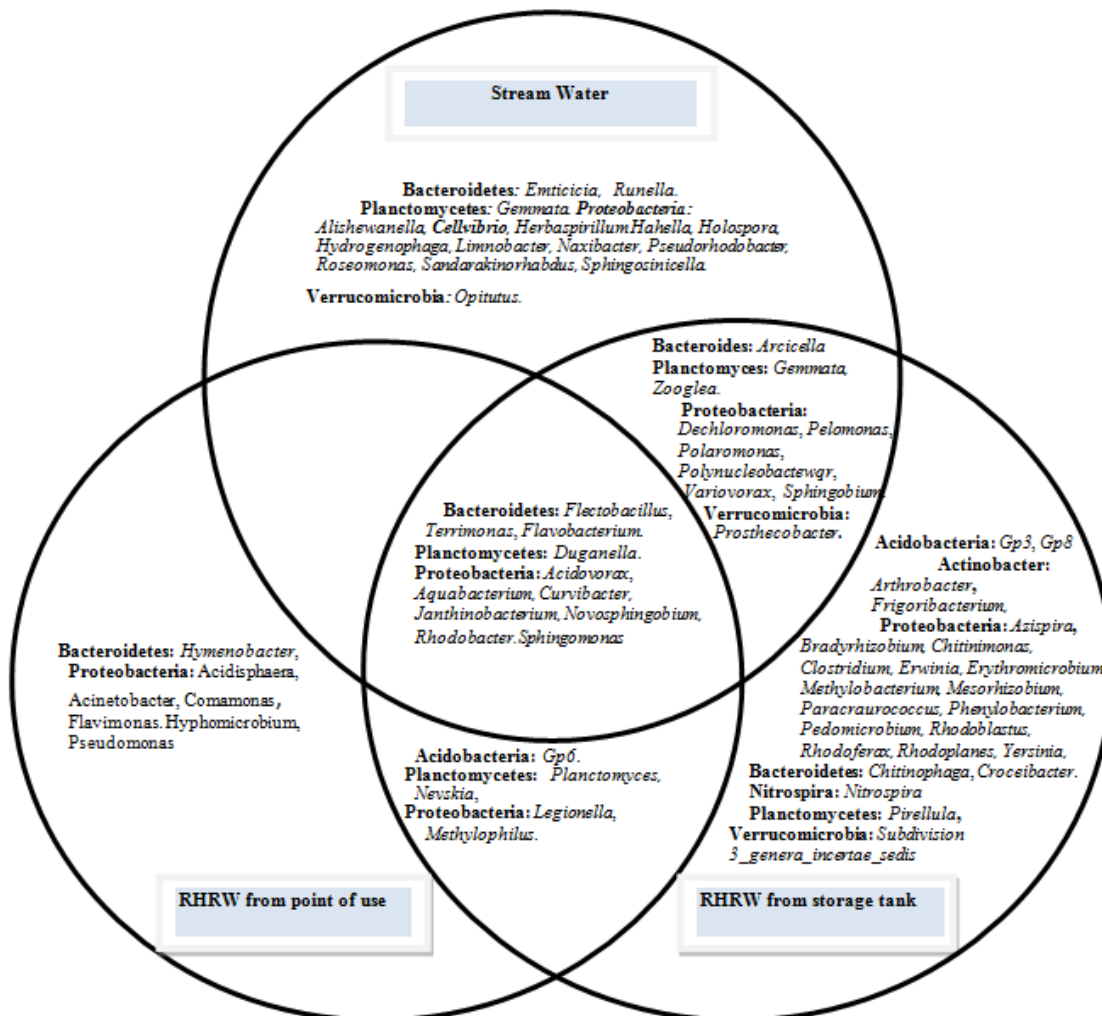


Figure 7. 7: Schematic drawing showing the detected bacterial genera found in the three different groups of water samples (Tank, River and point of use (kitchen)) in Luthegele village. Overlaps between the groups are indicated by the arrangement of the circles.

*Flectobacillus* (15.2%), and *Novosphingobium* (12.1%). In tank water (Figure 7.8C) *Curvibacter* (38.7%), *Gemmata* (12.9%) *Polynucleobacter* (11%) and *Terrimonas* (6%) were the most abundant genus. Uniquely distributed genera were *Acinebacter*, *Duganella*, *Methylophilus* detected only in kitchen water, *Cryobacterium* in tank 5, *Gp6*, *Nevskia* and *Planctomyces* in tank 1, *Hydrogenophaga*, *Cellvibrio*, *Opitutus*, *Prostheco bacter*, *Pseudorhodobacter*, *Roseomonas* and *Sphingobium* in river water and *Rhodoplans* in tank 7.

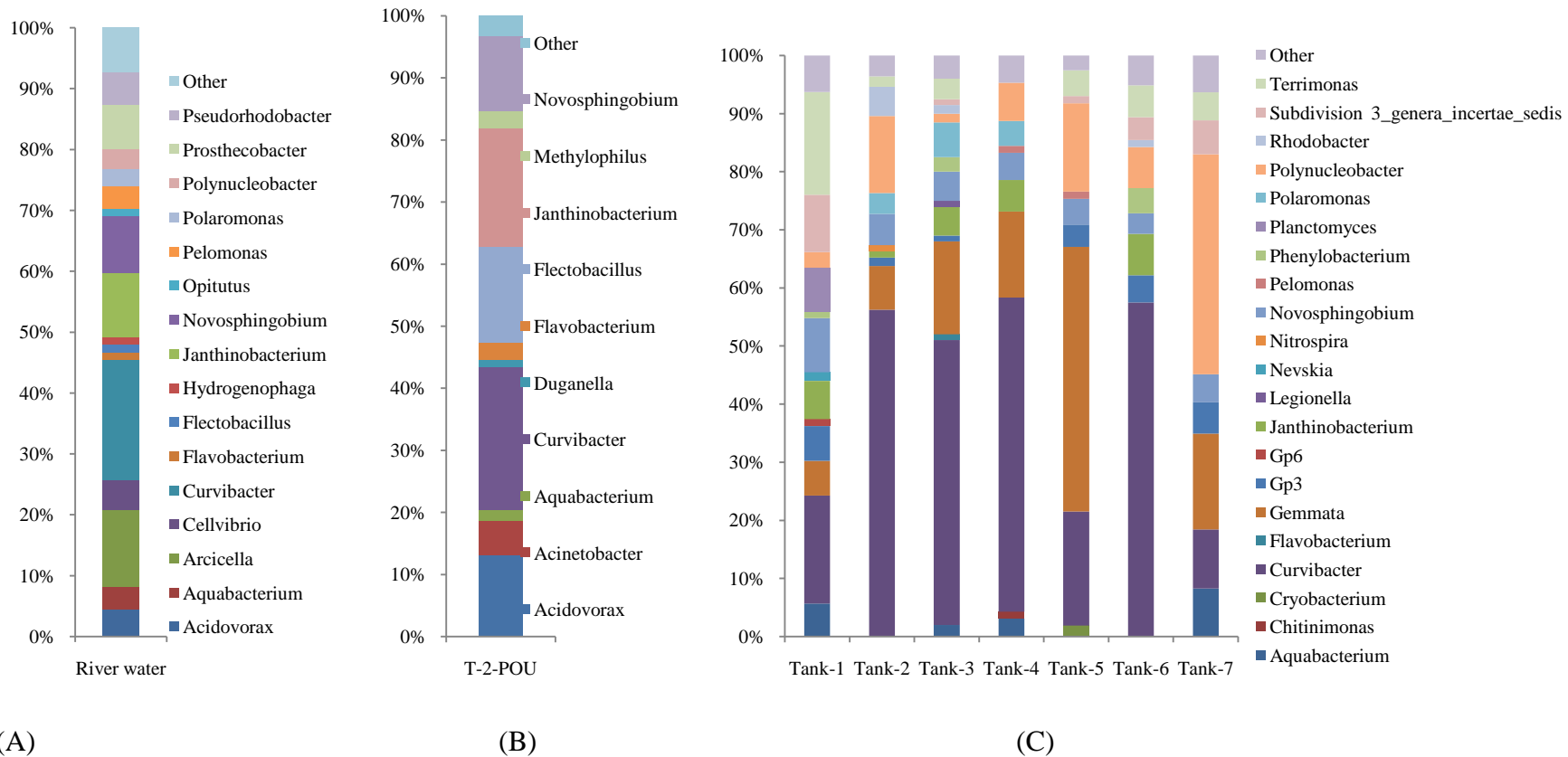


Figure 7. 8 Bacterial communities of river water (A), kitchen water (B) and tank water (C). Relative clone composition of genera was determined by pyrosequencing of 16S rRNA from metagenomic DNA extracted from the water collected from Luthengele village in the Eastern Cape Province. The identification 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.

### 7.4.3 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 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 Jackknifed weighted and unweighted UniFrac distances showed marked clustering differences (Figure 7.9). 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. These findings are similar to the results obtained with cluster analysis (Figure 7.6).

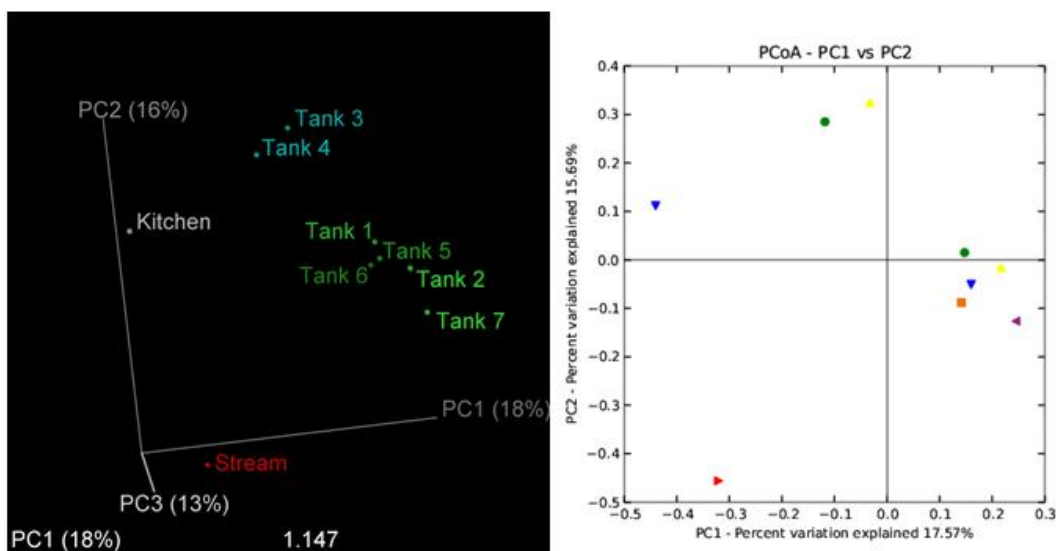


Figure 7. 9: Principal coordinate analysis of unweighted UniFrac distances. PCoA plots were made using jackknifed UniFrac distances in QIIME, (A) PC1 versus PC2, (B) PC1 versus PC3, (C) PC2 versus PC3.

### 7.4.4 Detection of pathogenic signatures

A total of eight genera that are known to contain pathogenic bacteria were considered (Table 7.2). 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 7.2.

Table 7. 2: Probable species\* of the detected genera known to contain pathogenic species (Tamura et al., 2011).

<b>Genera</b>	<b>Species</b>
<i>Serratia</i>	<i>nematodiphila, ureilytica</i>
<i>Tatlockia,</i>	<i>micdadei</i>
<i>Chromobacterium</i>	<i>aquaticum, haemolyticum, violaceum, subtsugae</i>
<i>Clostridium</i>	<i>algidicarnis, putrefaciens, sartagoforme, cellulovorans</i>
<i>Pseudomonas</i>	<i>Brassicacearum subsp. denitrificans, entomophila, putida, nitroreducens, fluorescens</i>
<i>Legionella</i>	<i>Pneumophila subsp. pascullei, pneumophila subsp. pneumophila, cincinnatiensis, longbeachae, fairfieldensis, adelaidensis, rubrilucens, taurinensis, gresilensis, uncultured Legionella_sp.</i>
<i>Yersinia</i>	<i>bercovieri, pestis, mollaretii, frederiksenii, ruckeri, massiliensis</i>
<i>Acinetobacter</i>	<i>oleivorans, calcoaceticus, haemolyticus, junii, parvus, gyllenbergii</i>

\*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 method. 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 *Acinetobacter 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 kitchen water, *Clostridia* (1 sequence) and *Chromobacterium* (1 sequence) in tank 1, *Yersinia* (2) in tank 1 and tank 5 and *Serratia* in in tank 4 and tank 7. *Legionella pneumophila* was the only species identified to be closely related to sequences from our pathogenic signatures, with respect to bacteria species of concern listed in water (US EPA, 2009).

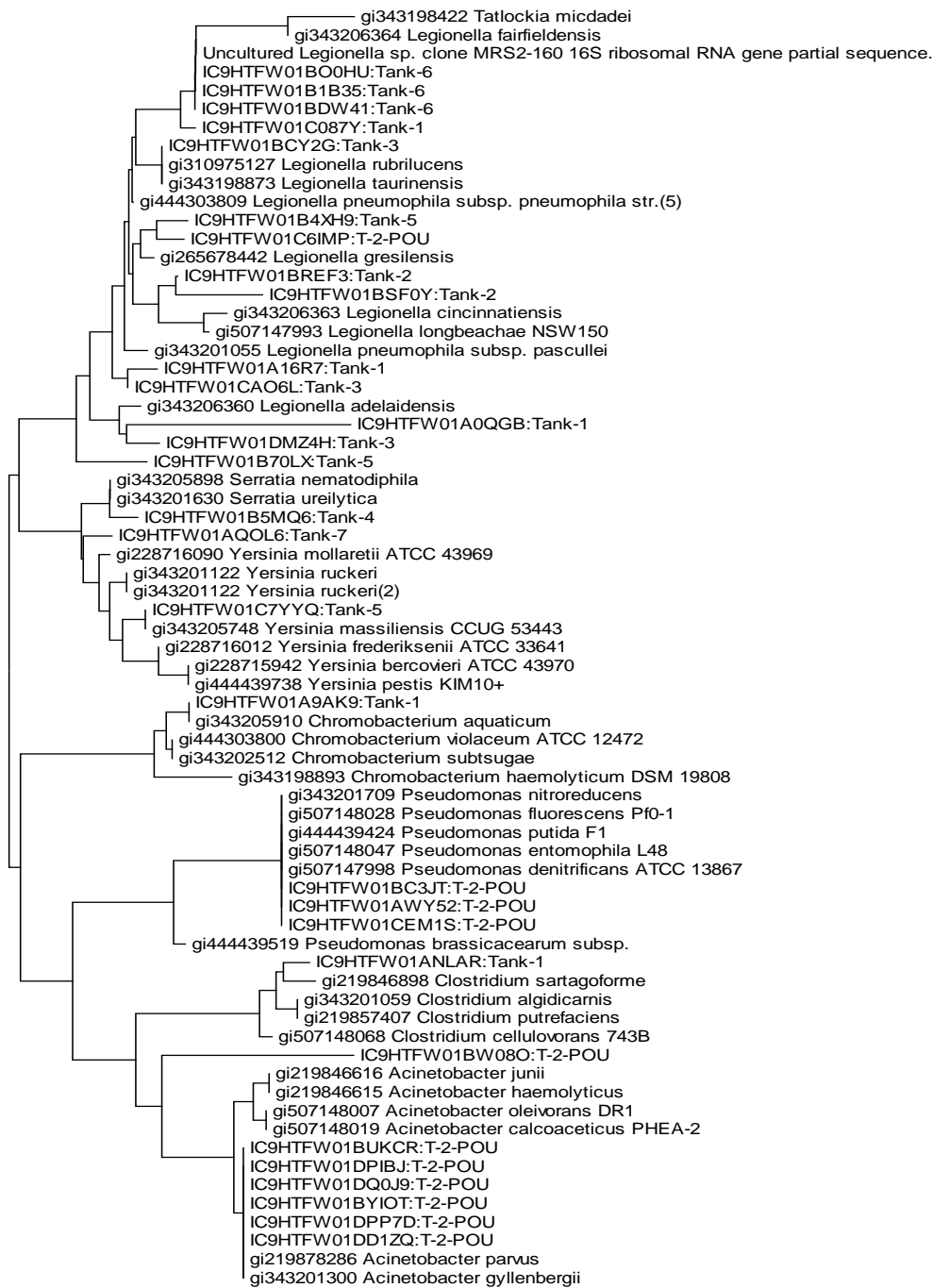


Figure 7. 10: Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The optimal tree where the sum of branch length of 2.07298780, is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Jukes-Cantor method (Jukes and Cantor, 1969) and are in the units of the number of base substitutions per site. The analysis involved sixty-nine nucleotide sequences. The first codon positions were included. All positions containing gaps and missing data were eliminated. There were a total of forty-five positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

## **7.5 Discussion**

### **7.5.1 Microbial diversity**

Little information exists about microbial communities in RHRW. In this study we used barcoded pyrosequencing of the V1-V3 hypervariable regions of the 16S rRNA 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 researchers 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 (Angus, 1983; Ashbolt, 2004).

Little information exists on the pyrosequencing study of microbial communities in RHRW and river water sources. The results of our study demonstrated diverse bacterial communities in RHRW, river and kitchen water samples. The 454 pyrosequencing analysis allowed for the detection of 1092 OTUs. 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 and the

potential of 454 pyrosequencing to detect rare organisms in microbial communities (Vaz-Moreira et al., 2011).

Of the eleven different phyla detected in water samples, phylotypes belonging to the phyla *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 phyla, *Verrucomicrobia* and *Plancomyces* have only been reported through culture-independent methods. *Firmicutes*, *Actinobacteria*, *Acidobacteria* and *OP10* were found in lower abundances as previously pointed out in other freshwater studies (Hong et al., 2011; 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 (Vaz-Moreira et al. 2011; Matcher et al. 2011; Steelman et al. 2012; Zwart et al. 2002), reverse line blot hybridisation (Lindström et al., 2005) or metagenomic and FISH analysis (Cottrell et al., 2005).

*Proteobacteria* was the most abundant and largest phylum in all the water samples. The two classes of this phylum, *Alphaproteobacteria* and *Betaproteobacteria* were present in relatively higher concentrations in all water samples, although the class *Alphaproteobacteria* was detected at relatively lower abundance ratios (Figure 7.4). The wide distribution of *Alphaproteobacteria* and *Betaproteobacteria* in freshwater environments has been well documented, and pH and nutrients have been reported to be related to their abundances (Newton et al., 2011). Previous studies have also reported phylotypes belonging to *Betaproteobacteria* to be dominant in freshwater systems, similar to our findings (Pinto et al., 2012). Pinto et al. (2012) also reported that a large proportion (45%) of the *Betaproteobacteria* phylotypes were unclassified bacteria. On the contrary only 21.3% *Betaproteobacteria* from our study could not be clarified to genus level.

*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. 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).

While this study did not evaluate the prevalence of different bacterial species in biofilms, it is well known that bacteria in fresh water systems including river and rainwater can survive and proliferate through biofilm formation (U. S. Environmental Protection Agency, 2002; Paulse et al., 2009; Bricheux et al., 2013). These biofilms are known to shed off in to the water, acting as a source of contamination (U. S. Environmental Protection Agency, 2002). 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 Proteobacteria have over other bacterial species.

### 7.5.2 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 phyla will depend on the immediate sources of contamination or the fresh water ecosystem (Dolan, 2005; 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 dirty 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 *Gamma-proteobacteria* 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.

### 7.5.3 Bacterial diversity

For a comparison of species richness among the nine samples belonging to three sample groups, rarefaction curves were generated at 3% (species level), 5% (genus level) and 20% (phylum level). The OTU rarefaction plot that 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. 5,000 reads per sample than the 855 –1,739 reads per sample generated in this study.

The number of 16S rRNA gene sequences analysed (10 956 - 18 510 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.

Most pyrosequencing studies on water microbial ecology deal with relatively small sample sizes. Hong et al. (2010) used pyrosequencing to compare bacterial communities in two water meters of a drinking water distribution system using individual samples of biofilm and water from the taps. Navarro-Noya et al. (2013) used pyrosequencing of nine single water samples to analysis the bacterial communities in nine drinking water wells. Orgiazzi et al. (2013) used pyrosequencing analysis of fungal assemblages in 16 individual soil samples from

geographically distant areas to reveal spatial patterning and a core mycobiome. The advantage with pyrosequencing is that although single samples are used, the technique sequences all the targeted microbes present. This results in thousands or tens of thousands of sequences representing individual bacteria. Consequently individual samples are still good enough for comparative purposes between samples from different environmental settings. Nevertheless multiple replicated samples will generally capture more diversity and provide a better overview on microbial communities.

#### **7.5.4 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, *Legionella pneumophila* is the only one included in the U. S. EPA list for 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* sp., 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).

The 16S rRNA gene sequences of these pathogenic genera were identified by comparing with sequences of known pathogens at a Jukes-Cantor distance of 0.03, which has been proposed



as a measure of species level similarity (Schloss and Handelsman, 2006). However, the challenge with *Legionella* contamination is their difficulty to control once they become established in a water system biofilm (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. that 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). However, The majority of non-pneumophila *Legionella* confirmed infections have been reported in immunosuppressed patients. The major challenge with *Legionella* species in water systems is their difficulty to remove once they have colonised a system (Kim et al., 2002).

Although the other probable species such as *Acinetobacter*, *Clostridium*, *Legionella*, *Pseudomonas*, *Serratia* and *Yersinia* detected in our study are not listed among the pathogens of concern, they can still cause significant infection especially in immune-compromised individuals. For example *Pseudomonas* are known to be associated with plants where they survive as symbionts. However the genus *Pseudomonas* has been reported also contain species that can undergo bivalent interactions with humans and may cause facultative infections in those with certain predispositions (Berg et al., 2005; Lai et al., 2011).

The sampling coverage and sequencing depth with respect to expected human exposure and risk, was not extensive enough. The total of 10956 16S rDNA sequences that were retrieved from the nine samples might not have been enough to detect pathogens which may be well underrepresented in the microbial communities. In a study by Ahmed et al. (2010b) on health risk from the use of RHRW in Southeast Queensland, Australia, quantitative PCR estimated numbers of *Salmonella*, *G. lamblia*, and *L. pneumophila* organisms ranged from  $6.5 \times 10^1$  to  $3.8 \times 10^2$  cells,  $0.6 \times 10^0$  to  $3.6 \times 10^0$  cysts, and  $6.0 \times 10^1$  to  $1.7 \times 10^2$  cells per 1,000 ml of water, respectively. Given these prevalence ratios and that pathogens usually constitute less than 1%

of the microbial population in most environments (Bibby et al., 2010), the required number of sequences per sample would need to run into millions per sample to accurately detect underrepresented pathogens in environmental waters.

Pathogenic bacteria of concern that have previously been identified in RHRW include *Clostridium*, *Bacteroides*, *Campylobacter*, *Legionella*, *Aeromonas*, *Salmonella* and *Yersinia* (Ahmed et al., 2008, 2010b). While we were able to detect sequences belonging to pathogenic genera including *Acinetobacter*, *Clostridium*, *Flavobacterium*, *Legionella*, *Methylobacterium*, *Pseudomonas*, *Serratia* 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  $6.9 \times 10^1$ – $3.8 \times 10^2$  per single event say one litre consumption may lead to a 1 in 10,000 risk, (Ahmed et al., 2010b). A  $10^2$  value, when compared to, say  $10^{10}$  total cells per litre 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 or sample pre-treatments to remove common sequences, such as DGGE 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. Furthermore, 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 7.10). 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.

## **7.6 Conclusion and recommendations**

The detection of pathogen concentrations in environmental samples including air, water, and wastewater are biased due to the required selection of individual or a limited groups of potential pathogens for analysis. The use of pyrosequencing technology combined with the ever-increasing sequencing depth and read lengths can potentially circumvent these biases by sequencing deeply enough into populations to describe the true diversity of pathogens and include both established and emerging agents (Bibby et al., 2010). The application of pyrosequencing technology used here to analyse RHRW samples and comparative controls enabled us to identify eight genera that are known to contain pathogenic species, although, we could not identify with certainty the species identities of the sequences in these genera (Table 7.2). 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. Currently the South African government is performing surveys of RHRW and alternative water sources in order to generate data that can be used to develop risk-based guidelines. The judicious implementation of pyrosequencing analysis can provide

crucial guidance for prioritizing culture-based and quantitative analysis of selected pathogens and ensure that significant pathogenic species are not left out of these risk assessments.

Although this study demonstrated complex microbial communities in river and RHRW sources of a rural community, it is not clear how the diversity of bacterial communities in these drinking water sources influences the quality of water or how the presence of pathogen signatures threatens the water safety. In addition, it is not clear why some potential pathogens e. g. *Legionella* spp. seem to be prevalent in RHRW storage in tanks. Further research is required to resolve these issues. Nevertheless, this study clearly demonstrated that numerous bacterial species inhabit both river and RHRW. Although RHRW is such a promising water source in the face of dwindling water resources, caution should be paid to potential risks. 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.

## 7.7 References

- Ahmed, W., A. Vieritz, A. Goonetilleke, and T. Gardner. 2010b. Health risk from potable and non-potable uses of roof harvested rainwater using quantitative microbial risk assessment in Australia. *Applied and Environmental Microbiology* 76: 7382–7391.
- Ahmed, W., F. Huygens, A. Goonetilleke, and T. Gardner. 2008. Real-Time PCR detection of pathogenic microorganisms in roof harvested rainwater in Southeast Queensland, Australia. *Applied and Environmental Microbiology* 74: 5490–5496.
- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Angus, K. W. 1983. Cryptosporidiosis in man, domestic animals and birds: A review. *Journal of the Royal Society of Medicine* 76: 62–70.
- Ashbolt, N. J. 2004. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology* 198: 229–238.

- Berg, G., L. Eberl, and A. Hartmann. 2005. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environmental Microbiology* 7: 1673–1685.
- Bibby, K., E. Viau, and J. Peccia. 2010. Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. *Water Research* 44: 4252–4260.
- Bricheux, G., L. Morin, G. Moal, G. Coffe, D. Balestrino, N. Charbonnel, J. Bohatier, and C. Forestier. 2013. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. *Microbiology Open* 2, 402-414.
- Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, and J. I. Gordon. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335–336.
- Chen, C. -L., W. T. Liu, M. L. Chong, M. T. Wong, S. L. Ong, H. Seah, and W. J. Ng. 2004. Community structure of microbial biofilms associated with membrane-based water purification processes as revealed using a polyphasic approach. *Applied Microbiology and Biotechnology* 63: 466–473.
- Claesson, M. J., O. O’Sullivan, Q. Wang, J. Nikkilä, J. R. Marchesi, H. Smidt, W. M. de Vos, R. P. Ross, and P. W. O’Toole. 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* 4: e6669.
- Cleary, D. F. R., K. Smalla, L. C. S. Mendonc, and N. C. M. Gomes. 2012. Assessment of variation in bacterial composition among microhabitats in a mangrove environment using DGGE fingerprints and barcoded pyrosequencing. *PLoS One* 7: 1–8.
- Costa, J., I. Tiago, M. S. da Costa, and A. Veríssimo. 2005. Presence and persistence of *Legionella* spp. in groundwater. *Applied and Environmental Microbiology* 71: 663–671.
- Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight. 2009. Bacterial community variation in human body habitats across space and time. *Science* 326: 1694–1697.
- Cottrell, M. T., L. A. Waidner, L. Yu, and D. L. Kirchman. 2005. Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environmental Microbiology* 7: 1883–1895.

- DeSantis, T. Z., E. L. Brodie, J. P. Moberg, I. X. Zubieta, Y. M. Piceno, and G. L. Andersen. 2007. High density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microbial Ecology* 53: 371–383.
- Dolan, J. R. 2005. Biogeography of aquatic microbes. *Aquatic Microbial Ecology* 41: 39–48.
- Dorfman, M. H., N. Stoner, and K. S. Rosselot. 2009. Testing the waters: A guide to water quality at vacation beaches. Natural Resources Defense Council, San Francisco.
- Eichler, S., R. Christen, C. Höltje, P. Westphal, J. Bötel, I. Brettar, A. Mehling, and M. G. Höfle. 2006. Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Applied and Environmental Microbiology* 72: 1858–1872.
- Fields, B. S., R. F. Benson, and R. E. Besser. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clinical Microbiology Reviews* 15: 506–526.
- Gich, F., K. Schubert, A. Bruns, H. Hoffelner, and J. Overmann. 2005. Specific detection, isolation, and characterisation of selected, previously uncultured members of the freshwater bacterioplankton community. *Applied and Environmental Microbiology* 71: 5908–5919.
- Hayashi, H., M. Sakamoto, and Y. Benno. 2001. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiology and Immunology* 46: 535–548.
- Hong, P. -Y., C. Hwang, F. Ling, G. L. Andersen, M. W. LeChevallier, W. Liu, L. Gary, and L. W. T. 2010. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Applied and Environmental Microbiology* 76: 5631–5635.
- Hur, I., and J. Chun. 2004. A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. *Journal of Microbiology* 42: 9–13.
- Huse, S. M., L. Dethlefsen, J. A. Huber, D. M. Welch, D. A. Relman, and M. L. Sogin. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS ONE* 4: e1000255.

- Janssen, P. H. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology* 72: 1719–1728.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules, Munro HN, Mammalian protein metabolism. Academic press, New York. pp. 21–132.
- Kahinda, J. M., A. E. Taigbenu, and R. J. Boroto. 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.
- Kahinda, J. M., E. S. B. Lillie, A. E. Taigbenu, M. Taute, and R. J. Boroto. 2008. Developing suitability maps for rainwater harvesting in South Africa. *Physics and Chemistry of the Earth Parts ABC* 33: 788–799.
- Kim, B. R., J. E. Anderson, S. A. Mueller, W. A. Gaines, and A. M. Kendall. 2002. Literature review—efficacy of various disinfectants against *Legionella* in water systems. *Water Research* 36: 4433–4444.
- Lai, C. -C., A. Cheng, W. -L. Liu, C. -K. Tan, Y. -T. Huang, K. -P. Chung, M. -R. Lee, and P. -R. Hsueh. 2011. Infections caused by unusual *Methylobacterium* species. *Journal of Clinical Microbiology* 49: 3329–3331.
- Lautenschlager, K., C. Hwang, W. Liu, N. Boon, O. Ko, H. Vrouwenvelder, T. Egli, and F. Hammes. 2013. A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. *Water Research* 47: 3015–3025.
- Lindström, E. S., M. P. Kamst-Van Agterveld, and G. Zwart. 2005. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Applied and Environmental Microbiology* 71: 8201–8206.
- Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight. 2007. Quantitative and qualitative  $\beta$  diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* 73: 1576–1585.
- Lozupone, C., M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight. 2010. UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal* 5: 169–172.

- Luxmy, B., F. Nakajima, and K. Yamamoto. 2000. Analysis of bacterial community in membrane-separation bioreactors by fluorescent in situ hybridisation (FISH) and denaturing gradient gel electrophoresis (DGGE) techniques. *Water Science and 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: 5429–5434.
- MacLean, D., J. D. G. Jones, and D. J. Studholme. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nature Reviews Microbiology* 7: 287–296.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. -J. Chen, Z. Chen, Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. -J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. -B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., J. M. Rothberg. 2005. Genome sequencing in microfabricated high density picolitre reactors. *Nature* 437: 376–380.
- Martiny, A. C., H. -J. Albrechtsen, E. Arvin, and S. Molin. 2005. Identification of bacteria in biofilm and bulk water samples from a nonchlorinated model drinking water distribution system: Detection of a large nitrite-oxidizing population associated with *Nitrospira* spp. *Applied and Environmental Microbiology* 71: 8611–8617.
- Matcher, G. F., R. A. Dorrington, T. O. Henninger, and P. W. Froneman. 2011. Insights into the bacterial diversity in a freshwater-deprived permanently open Eastern Cape estuary, using 16S rRNA pyrosequencing analysis. *Water SA* 37: 381–390.
- Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder. 2004. Source tracking fecal bacteria in water: A critical review of current methods. *Journal of Environmental Management* 73: 71–79.



- Metzker, M. L. 2009. Sequencing technologies—the next generation. *Nature Reviews Genetics* 11: 31–46.
- Muder, R. R., and V. L. Yu. 2002. Infection due to *Legionella* species other than *L. pneumophila*. *Clinical Infectious Diseases* 35: 990–998.
- Murga, R., T. S. Forster, E. Brown, J. M. Pruckler, B. S. Fields, and R. M. Donlan. 2001. Role of biofilms in the survival of *Legionella pneumophila* in a model potable water system. *Microbiology* 147: 3121–3126.
- Navarro-Noya, Y. E., M. C. Suárez-arriaga, A. Rojas-Valdes, N. M. Montoya-Ciriaco, S. Gómez-Acata, F. Fernández-Luqueño, and L. Dendooven. 2013. Pyrosequencing analysis of the bacterial community in drinking water wells. *Microbial ecology* 66: 19–29.
- Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews* 75: 14–49.
- Nocker, A., M. Burr, and A. K. Camper. 2007. Genotypic microbial community profiling: A critical technical review. *Microbial Ecology* 54: 276–289.
- Orgiazzi, A., V. Bianciotto, P. Bonfante, S. Daghino, S. Ghignone, A. Vizzini, S. Bagella, C. Murat, and M. Girlanda. 2013. Pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. *Diversity* 5: 73–98.
- Palmer, C. J., Y. L. Tsai, C. Paszko-Kolva, C. Mayer, and L. R. Sangermano. 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: 3618–3624.
- Paulse, A. N., V. A. Jackson, and W. Khan. 2009. Comparison of microbial contamination at various sites along the Plankenburg- and Diep Rivers, Western Cape, South Africa. *WaterSA* 35: 469–478.
- Pinto, A. J., C. Xi, and L. Raskin. 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environmental Science and Technology* 46: 8851–8859.

- Poitelon, J. -B., M. Joyeux, B. Welté, J. -P. Duguet, E. Prestel, O. Lespinet, and M. S. DuBow. 2009. Assessment of phylogenetic diversity of bacterial microflora in drinking water using serial analysis of ribosomal sequence tags. *Water Research* 43: 4197–4206.
- Rasid, R. A. 2009. Biofilm and multimedia filtration for rainwater treatment. *Journal of Sustainable Development* 2: 196–199.
- Roesch, L. F. W., R. R. Fulthorpe, A. Riva, G. Casella, A. K. M. Hadwin, A. D. Kent, S. H. Daroub, F. A. O. Camargo, W. G. Farmerie, and E. W. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal* 1: 283–290.
- Rowbotham, T. J. 1986. Current views on the relationships between amoebae, legionellae and man. *Israel Journal of Medical Sciences* 22: 678.
- Rusin, P. A., J. B. Rose, C. N. Haas, and C. P. Gerba. 1997. Risk assessment of opportunistic bacterial pathogens in drinking water. In *Reviews of environmental contamination and toxicology* (pp. 57-83). Springer, New York.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.
- Sanapareddy, N., T. J. Hamp, L. C. Gonzalez, H. A. Hilger, A. A. Fodor, and S. M. Clinton. 2009. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. *Applied and Environmental Microbiology* 75: 1688–1696.
- Schlesner, H., C. Jenkins, and J. T. Staley. 2006. The phylum Verrucomicrobia: A phylogenetically heterogeneous bacterial group. In *The Prokaryotes* (pp. 881–896). Springer, New York.
- Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* 71: 1501–1506.
- Schloss, P. D., and J. Handelsman. 2006. Towards a census of bacteria in soil. *PLoS Computational Biology* 2: e92.
- Sinclair, R., S. A. Boone, D. Greenberg, P. Keim, and C. P. Gerba. 2008. Persistence of category “A” select agents in the environment. *Applied and Environmental Microbiology* 74: 555–563.

- Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences* 103: 12115–12120.
- Stelman, S. M., B. P. Chowdhary, S. Dowd, J. Suchodolski, and J. E. Janecka. 2012. Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Veterinary Research* 8: 1–11.
- Stocker, S., R. Snajder, J. Rainer, S. Trajanoski, G. Gorkiewicz, Z. Trajanoski, and G. G. Thallinger. 2011. SnoWMA: High-throughput phylotyping, analysis and comparison of microbial communities. *Proceedings of the 110<sup>th</sup> ASM General Meeting* 23 May 2010.
- Stout, J. E., V. L. Yu, and M. G. Best. 1985. Ecology of *Legionella pneumophila* within water distribution systems. *Applied and Environmental Microbiology* 49: 221–228.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- Tringe, S. G., and P. Hugenholtz. 2008. A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Biotechnology* 11: 442–446.
- U. S. Environmental Protection Agency. 2002. Health risks from microbial growth and biofilms in drinking water distribution systems. U. S. Environmental Protection Agency Office of Ground Water and Drinking Water Standards and Risk Management, Pennsylvania, Washington DC.
- US EPA. 2009. Drinking water contaminant candidate List 3. US EPA. *Federal Register* 74: 51850–51862.
- Vaz-Moreira, I., C. Egas, O. C. Nunes, and C. M. Manaia. 2011. Culture-dependent and culture-independent diversity surveys target different bacteria: A case study in a freshwater sample. *Antonie Van Leeuwenhoek* 100: 245–257.

- Wang, Y., and P. -Y. Qian. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4: e7401.
- Ward, N., J. T. Staley, J. A. Fuerst, S. Giovannoni, H. Schlesner, and E. Stackebrandt. 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* (pp757–793). Springer, New York.
- World Health Organisation. 2004. Guidelines for drinking water quality: recommendations. World Health Organisation, Geneva
- Zwart, G., B. C. Crump, M. P. Kamst-van Agterveld, F. Hagen, and S. -K. Han. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology* 28: 141–155.

## CHAPTER EIGHT

### GENERAL DISCUSSION

The use of roof harvested rainwater (RHRW) is currently increasing in South Africa, with both government and non-governmental organisations promoting the practice (Kahinda et al., 2007). Rainfall harvested from rooftops has been promoted for portable use, while rainfall collected from surface runoff has been promoted for domestic food gardening (Denison and Wotshela, 2009). Research on rainwater harvesting (RWH) in rural settings such as the ones in South Africa is limited and data from other situations cannot be directly extrapolated to rural settings (Dobrowsky et al., 2014; Gómez-Duarte et al., 2010). Given the lack of standards, knowledge of contamination and complexity of the problem with RWH, there is a need to propose a set of guidance information on best practices and risks associated with harvesting rainwater. This guidance information more specifically should detail the targeted water quality range in which infection or adverse effects are unlikely to occur.

Findings from our studies revealed significant variations in water quality between samples from various environmental settings. A variety of makeshift containers are used in RWH systems in rural communities and neither first flush diverter nor filtration system are employed in most cases. Comparatively households that use modern RWH systems had first flush diverters and a filtration system installed. Consequently the samples from modern RWH systems had the lowest level microbial concentrations detected. While the use of modern rainwater systems would be ideal in terms of guaranteeing water quality, the installation of such system would be too expensive for rural households who can hardly afford the cost of purchasing a RWH tank.

In Luthengele village in the Eastern Cape Province of South Africa households use river water, RHRW and spring water for potable purposes without prior treatment. Analysis of the microbiological quality of river water did not show it to be worse off than roof harvested rainwater in the observed microbiological counts. Since these rivers are constantly flowing, from the ground water table, dilution effect play a significant role in the microbiological counts observed. While the observed microbiological counts may suggest the river water to

be relatively clean, river water is an open system that is subject to many uncertain factors of contamination.

Water in South African rural communities is fetched from the source (tank, river, borehole, etc) using a variety of containers and stored prior to use. Although the water might be clean at the source, the process of fetching and storing prior to use exposes it to potential contamination. Comparison of the microbiological quality of RHRW sampled directly from the tanks and that which had been stored temporarily prior to use showed significant variations in their microbiological quality. The findings suggest that containers used to store the water or the process of fetching it, to be responsible for the contamination detected. It is most probable that bacteria may accumulate in the containers in the form of biofilm creating a perennial source of contamination (Camper et al., 1998).

Heterotrophic plate counts, faecal coliforms, *E. coli*, and enterococci are the traditionally used bacteria of choice, of which *E. coli* is the most used and recognised indicator bacteria (Edberg et al., 2000). However, our findings do not suggest *E. coli* to be a suitable indicator in RHRW systems, as has been previously reported (Ahmed et al., 2011). In most of the tanks from which *E. coli* were detected, the concentrations were up to ten times lower than those for enterococci concentrations. Consequently, the suitability of *E. coli* as indicator bacteria in harvested rainwater systems has been questioned (Spinks et al., 2006; Ahmed et al., 2011). The higher prevalence of enterococci in RHRW may be a result of their better capacity to persist in water longer than *E. coli* (McFeters et al., 1974). A number of researches in environmental water quality have also reported enterococci to perform better in faecal contamination indication compared to *E. coli* (Savichtcheva and Okabe, 2007; Schets et al., 2010).

Evaluation of the prevalence of enterococci species showed the presence of *E. casseliflavus*, *E. faecalis*, *E. faecium*, *E. hirae* and *E. mundtii* in both RHRW and pigeon faecal samples. The prevalence of *E. faecalis* (20.5%), *E. mundtii* (20.5%), *E. faecium* (23.1%) and *E. casseliflavus* (17.3%) in bird faecal samples suggests that the birds may have contributed to the enterococci detected in rainwater tanks. It is also possible that other sources of

enterococci not tested in this study, including dust, vegetation and animals such as lizards, rats or frogs may also have contributed to the enterococci detected in RHRW tanks (Evans et al., 2006a; Ahmed et al., 2012b). These findings show the significance of enterococci speciation in its use as indicator bacteria of faecal contamination. Furthermore, the lack of one dominant enterococci species in both faecal and RHRW samples suggests that no single enterococci is a reliable indicator of the host faecal source. However, strain diversity characterisation of enterococci have been suggested to provide supporting evidence for bacterial source identification whether it be faecal, vegetation or soil (Layton et al., 2010).

The mitigation of environmental water faecal pollution/contamination requires rapid species identification and strain diversity determination of specific bacteria species which then can be used to identify potential sources of contamination, generally termed bacterial source tracking (Meays et al., 2004). In this study we applied MALDI-TOF-MS for the characterisation and differentiation of *E. coli* strains isolated from RHRW and pigeon faeces, river water, sewage water, bovine milk and humans, in an attempt to evaluate the applicability of MALDI-TOF-MS in distinguishing *E. coli* strains and possible application in bacterial source tracking (BST). An overall mean similarity between duplicate MALDI-TOF-MS profiles of  $96.85 \pm 2.10\%$  was achieved. These results imply good reproducibility and are in agreement with results from a previous study (Mellmann et al., 2009). Analysis of data in ClinProtTools software, using genetic algorithm (GA) and artificial neural network (ANN) analysis schemes showed them to be slow and were further complicated by the large dataset. The use of more complicated statistical algorithms including support vector machine (SVM) brought improvement of algorithms resulting in a high recognition rate (Cho et al., 2012).

SVM data analyses were performed with the aim of evaluating the discriminatory effect in the data set. This kind of analyses gives an indication on how well the technique can discriminate between the MALDI-TOF mass spectra of the different strains. The application of SVMs to the binary peak data and the test results showed a very high strain identification performance. MALDI-TOF-MS analysis has been shown to have a very high identification capacity among *E. coli* species and SVM take advantage of the peak patterns embedded in the MALDI-TOF mass spectra. This SVM approach therefore enables the extension of MALDI-TOF-MS spectra in bacterial characterisation.

To help evaluate the potential health risk associated with the use of RHRW and the significance of pigeon faeces as the most likely major source of contamination, we characterised antibiotic resistance profiles in *E. coli* isolated from RHRW tanks and urban pigeon faeces. The motivation for this was the fact that previous studies on RHRW have implicated droppings of birds, mammals and/or rodents that have access to catchment areas or water storage tanks as the possible sources of microbial contamination (Martin et al., 2010; Amin and Ham, 2009; Ahmed et al., 2008; Dillaha and Zolan, 1985). Of these animals, birds are the most mobile and have been implicated as reservoirs of antibiotic resistant enteric pathogens (Silva et al., 2009). The presence of pigeon faeces in urban environments may contribute to the spread of infectious agents through the dispersion of wind-dried dropping particles (Oliveira and Pinhata, 2008; Marques et al., 2007).

The most frequently encountered form of resistance in all samples was to ampicillin (26.9%), gentamicin (26%), amoxicillin (25.2%), tetracycline (18.2%), and amikacin (17.8%) (Table 6.2). Most of these antibiotics have been widely used for therapeutic purposes against bacterial infections in humans and animals as well as growth promoters in agriculture and aquaculture (Kruse and Sorum, 1994; Khachatourians, 1998). The observation of relatively high antibiotic resistance percentages to similar antibiotics for *E. coli* from RHRW and birds, suggest birds to be the source of RHRW contamination.

Isolates from birds had a relatively higher proportion of antibiotic resistance than those from RHRW tanks, with the most significant differences being noted in resistance to gentamicin (12.3%), tetracycline (9.8%), chloramphenicol (6.5%), and amikacin (4.2%) suggesting *E. coli* sources other than bird droppings only. Hence it can be suggested that there are other sources of faecal contamination at play, which harbour *E. coli* with less AR prevalence. Probable sources of these *E. coli* are natural environmental habitats, other species of birds, small animals such as rats and dust particles with *E. coli* populations that are not as exposed to antibiotic as are other domestic animals.



Similar to our findings, previous research have shown *E. coli* strains isolated from pigeon faeces to be antibiotic resistant and may reflect the abusive use of such substances in our society (Silva et al., 2009). Pigeons possess a less developed cecum, which is the part of the gastrointestinal tract that harbours the most abundant and diversified microbiota. Hence, it is possible that these microorganisms may not be true inhabitants of pigeon faecal flora. Consequently, it is less likely that these birds could have a permanent gut microbiota with intestinal bacterial species commonly associated with humans (Baele et al., 2002). However, the feeding habits of urban pigeons may include garbage from nearby trash cans, which exposes them to contamination with residual antimicrobials or chemicals and medically important bacteria (Haag-Wackernagel and Moch, 2004; Rosengren et al., 2009).

Despite the lack of reports on antimicrobial susceptibility patterns of *E. coli* from RHRW and their comparison to the likely sources of faecal origin, a number of researches have shown increasing resistance to tetracycline, ampicillin and gentamicin, with resistance at lower levels to cotrimoxazole, nalidixic acid chloramphenicol enrofloxacin cefoxitin and ciprofloxacin, similar to our findings (Cardonha et al., 2005; Tanaka et al., 2005; Kahlmeter, 2003).

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 and river water used for domestic purposes in Luthengele village, Eastern Cape Province of South Africa. The results of our study demonstrated diverse bacterial communities in RHRW, river and kitchen water samples. A prominent feature of our analysis was the presence of unclassified bacteria. These findings suggest the existence of diverse bacteria that characterise the microbial community structure in the water and the potential of 454 pyrosequencing to detect rare organisms in microbial communities (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 in a number of studies (Gich et al., 2005; Matcher et al., 2011; Vaz-Moreira et al., 2011; Cleary et al., 2012; Steelman et al., 2012). *Proteobacteria* was the most abundant and largest phylum in all the water samples and the two classes of this phylum, *Alphaproteobacteria* and *Betaproteobacteria* were present in relatively higher concentrations in all water samples, although the class *Alphaproteobacteria* was detected at relatively lower abundance ratios. The wide distribution of *Alphaproteobacteria* and *Betaproteobacteria* in freshwater environments has been well documented and pH and nutrients have been reported to be related to their abundances (Newton et al., 2011). Previous studies have also reported phylotypes belonging to *Betaproteobacteria* to be dominant in freshwater systems, similar to our findings (Pinto et al., 2012). Pinto et al. (2012) also reported that a large proportion (45%) of the *Betaproteobacteria* phylotypes were unclassified bacteria. On the contrary only 21.3% *Betaproteobacteria* from our study could not be clarified to genus level.

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 (Dolan, 2005; 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. RHRW 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.

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. Of the

species that were closely related to our detected pathogenic signatures for the genera, *Acinetobacter*, *Clostridium*, *Legionella*, *Pseudomonas*, *Serratia* and *Yersinia*, only the species *Legionella pneumophila* is included in the U. S. Environmental Protection Agency 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). Approximately half of the *Legionella* species have been associated with disease in humans, with *Legionella pneumophila* infection being responsible for the majority of illnesses. Pathogenic bacteria of concern that have previously been identified in RHRW include *Clostridium*, *Bacteroides*, *Campylobacter*, *Legionella*, *Aeromonas*, *Salmonella* and *Yersinia* (Ahmed et al., 2008; 2010b). While we were able to detect sequences belonging to pathogenic genera including *Acinetobacter*, *Clostridium*, *Flavobacterium*, *Legionella*, *Methylobacterium*, *Pseudomonas*, *Serratia* and *Yersinia* their values were low and more sequencing depth is required to either exclude them or more confidently describe their presence.

## References

- Ahmed, W., A. Vieritz, A. Goonetilleke, and T. Gardner. 2010. Health risk from potable and non-potable uses of roof harvested rainwater using quantitative microbial risk assessment in Australia. *Applied and Environmental Microbiology* 76: 7382–7391.
- Ahmed, W., F. Huygens, A. Goonetilleke, and T. Gardner. 2008. Real-Time PCR detection of pathogenic microorganisms in roof-harvested rainwater in Southeast Queensland, Australia. *American Society for Microbiology (ASM). Applied and Environmental Microbiology* 74: 5490–5496.
- Ahmed, W., J. P. S. Sidhu, and S. Toze. 2012. Speciation and frequency of virulence genes of enterococcus spp. isolated from rainwater tank samples in Southeast Queensland, Australia. *Environmental Science Technology* 46: 6843–6850.
- Ahmed, W., T. Gardner, and S. Toze. 2011. Microbiological quality of roof-harvested rainwater and health risks: A review. *Journal of Environmental Quality* 40: 13–21.
- Amin, M. T. and M. Y. Ham. 2009. Roof harvested rainwater for potable purposes: application of solar collector disinfection (SOCO-DIS). *Water Research* 43: 5225–5235.

- Baele, M., L. A. Devriese, P. Butaye and F. Haesebrouck. 2002. Composition of enterococcal and streptococcal flora from pigeon intestines. *Journal of Applied Microbiology* 92: 348–351.
- Bricheux, G., L. Morin, G. Moal, G. Coffe, D. Balestrino, N. Charbonnel, J. Bohatier, and C. Forestier. 2013. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. *Microbiology Open* 2: 402–414.
- Camper, A. M. Burr, B. Ellis, P. Butterfield, and C. Abernathy. 1998. Development and structure of drinking water biofilms and techniques for their study. *Journal of Applied Microbiology* 85: 1S–12S.
- Cardonha, A. M., R. H. Vieira, G. Peirano and P. Rodrigues. 2005. Resistance to antibiotics and heavy metals from *Escherichia coli* isolated from sea water and pluvial galleries. *Journal Acta Cirurgica Brasileira* 20: 253–256.
- Chen, C. L., W. T. Liu, M. L. Chong, M. T. Wong, S. L. Ong, H. Seah, and W. J. Ng. 2004. Community structure of microbial biofilms associated with membrane-based water purification processes as revealed using a polyphasic approach. *Applied Microbiology and Biotechnology* 63: 466–473.
- Cho, Y., Y. Chiang, J. Shiea, and M. Hou. 2012. Combining MALDI-TOF and molecular imaging with principal component analysis for biomarker discovery and clinical diagnosis of cancer. *Genomic Medicine, Biomarkers and Health Sciences* 4: 3–6.
- Cleary, D. F. R., K. Smalla, L. C. S. Mendonc, and N. C. M. Gomes. 2012. Assessment of Variation in bacterial composition among microhabitats in a mangrove environment using DGGE fingerprints and barcoded pyrosequencing. *PloS One* 7: 1–8.
- Costa, J., I. Tiago, M. S. da Costa, and A. Veríssimo. 2005. Presence and persistence of *Legionella* spp. in groundwater. *Applied and Environmental Microbiology* 71: 663–671.
- Denison, J., and L. Wotshela. 2009. Indigenous water harvesting and conservation practices : Historical context, cases and implications. WRC Report No. TT 392/09. Water Research Commission (WRC). Pretoria.
- Dobrowsky, P. H., A van Deventer, M. De Kwaadsteniet, T. Ndlovu, S. Khan, T. E. Cloete, and W. Khan. 2014. Prevalence of virulence genes associated with pathogenic

- Escherichia coli* strains isolated from domestically harvested rainwater during low- and high-rainfall periods. *Applied and Environmental Microbiology* 80: 1633–1638.
- Dolan, J. R. 2005. Biogeography of aquatic microbes. *Aquatic Microbial Ecology* 41: 39–48.
- Edberg, S. C., E. W. Rice, R. J. Karlin, and M. J. Allen. 2000. *Escherichia coli*: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology* 88: 106S–116S.
- Eichler, S., R. Christen, C. Hölzje, P. Westphal, J. Bötel, I. Brettar, A. Mehling, and M. G. Höfle. 2006. Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Applied and Environmental Microbiology* 72: 1858–1872.
- Evans, C. A., P. J. Coombes, and R. H. Dunstan. 2006. Wind, rain and bacteria: The effect of weather on the microbial composition of roof-harvested rainwater. *Water Research* 40: 37–44.
- Gich, F., K. Schubert, A. Bruns, H. Hoffelner, and J. Overmann. 2005. Specific detection, isolation, and characterisation of selected, previously uncultured members of the freshwater bacterioplankton community. *Applied and Environmental Microbiology* 71: 5908–5919.
- Gómez-Duarte, O. G., J. Bai, and E. Newel. 2010. Detection of *Escherichia coli*, *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica*, *Vibrio cholerae*, and *Campylobacter* spp. enteropathogens by Three-reaction Multiplex PCR. *Diagnostic Microbiology and Infectious Disease* 63: 1–9.
- Haag-Wackernagel, D., and H. Moch. 2004. Health hazards posed by feral pigeons. *Journal of Infection*. 48: 307–313.
- Silva, V. L., Nicoli, J. R., Nascimento, T. C., and Diniz, C. G. . 2009. Diarrheagenic *Escherichia coli* Strains recovered from urban pigeons (*Columba livia*) in Brazil and their antimicrobial susceptibility patterns. *Current Microbiology* 59: 302–308.
- Kahinda, J. M., A. E. Taigbenu, and J. R. Boroto. 2007. Domestic rainwater harvesting to improve water supply in rural South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 32: 1050–1057.

- Kahlmeter, G. 2003. An international survey of the antimicrobial susceptibility of pathogens from uncomplicated urinary tract infections: The ECO. SENS Project. *Journal of Antimicrobial Chemotherapy*. 51: 69–73.
- Khachatourians, G. G. 1998. Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *Canadian Medical Association Journal* 159: 1129-1136.
- Kruse, H. and Sorum, H. 1994. Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Applied and Environmental Microbiology* 60: 4015-4021.
- Layton, B. A., S. P. Walters, L. H. Lam, and A. B. Boehm. 2010. *Enterococcus* species distribution among human and animal hosts using multiplex PCR. *Journal of Applied Microbiology* 109: 539–547.
- Martin, A. R., P. J. Coombes, T. L. Harrison and R. H. Dunstan. 2010. Changes in abundance of heterotrophic and coliform bacteria resident in stored water bodies in relation to incoming bacterial loads following rain events. *Journal of Environmental Monitoring*. 12: 255–260.
- Matcher, G. F., R. A. Dorrington, T. O. Henninger, and P. W. Froneman. 2011. Insights into the bacterial diversity in a freshwater-deprived permanently open Eastern Cape estuary, using 16S rRNA pyrosequencing analysis. *Water SA* 37: 381–390.
- McFeters, G. A., G. K. Bissonnette, J. J. Jezeski, C. A. Thomson, and D. G. Stuart. 1974. Comparative survival of indicator bacteria and enteric pathogens in well water. *Journal of Applied Microbiology* 27: 823–829.
- Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder. 2004. Source tracking fecal bacteria in water: a critical review of current methods. *Journal of Environmental Management* 73: 71–79.
- Mellmann, A., F. Bimet, C. Bizet, A. D. Borovskaya, R. R. Drake, U. Eigner, A. M. Fahr, Y. He, E. N. Ilina and M. Kostrzewa, 2009. High interlaboratory reproducibility of matrix-assisted laser desorption ionisation-time of flight mass spectrometry-based species identification of nonfermenting bacteria. *Journal of Clinical Microbiology* 47: 3732–3734.

- Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews* 75: 14–49.
- Oliveira, A. J. F. C., and J. M. W. Pinhata. 2008. Antimicrobial resistance and species composition of enterococci isolated from waters and sands of marine recreational beaches in South eastern Brazil. *Water Research*. 42: 2242–2250. Marques, S. M. T., R. M. de Cuadros, C. J. Silva, and M. Baldo. 2007. Parasites of pigeons (*Columba livia*) in urban areas of lages, Southern Brazil. *Parasitología latinoamericana*. 62: 183–187.
- Palmer, C. J., Y. L. Tsai, C. Paszko-Kolva, C. Mayer, and L. R. Sangermano. 1993. Detection of *Legionella* species in sewage and ocean water by polymerase chain reaction, direct fluorescent-antibody, and plate culture methods. *Journal of Applied and Environmental Microbiology* 59: 3618–3624.
- Pinto, A. J., C. Xi, and L. Raskin. 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environmental Science and Technology* 46: 8851–8859.
- Rosengren, L. B., C. L. Waldner and R. J. Reid-Smith. 2009. Associations between antimicrobial resistance phenotypes, antimicrobial resistance genes and virulence genes of faecal *Escherichia coli* from healthy grow-finish pigs. *Applied and Environmental Microbiology* 75: 1373– 1380.
- Savichtcheva, O. and S. Okabe. 2007. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application. *Water Research* 40: 2463–2476.
- Schets, F. M., R. Italiaander, H. H. J. L. Van Den Berg, and A. M. De Roda Husman. 2010. Rainwater harvesting: quality assessment and utilisation in The Netherlands. *Journal of Water And Health* 8: 224–235.
- Spinks, J., S. Phillips, P. Robinson, and P. Buynder. 2006. Bushfires and tank rainwater quality: A cause for concern *Journal of Water Health* 4: 21–28.
- Steelman, S. M., B. P. Chowdhary, S. Dowd, J. Suchodolski, and J. E. Jane. 2012. Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Veterinary Research* 8: 1–11.

- Tanaka, C., T. Miyazawa, M. Watari and N. Ishiguro. 2005. Bacteriological survey of faeces from feral pigeons in Japan. *Journal Veterinary Medical Science*. 67: 951–953.
- U. S. Environmental Protection Agency. 2009. Drinking Water Contaminant Candidate List 3. US EPA. Federal Register 74: 51850–51862.
- Vaz-Moreira, I., C. Egas, O. C. Nunes, and C. M. Manaia. 2011. Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample. *Springer. Antonie Van Leeuwenhoek* 100: 245–257.
- Zwart, G., B. C. Crump, M. P. Kamst-van Agterveld, F. Hagen, and S. -K. Han. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Inter-Research. Aquatic Microbial Ecology* 28: 141–155.