METAGENOMIC ANALYSIS OF BACTERIAL COMMUNITIES IN DRINKING WATER DISTRIBUTION SYSTEMS IN MOMBASA COUNTY (KENYA)

MWAJUMA JUNE JEOPHITA
I84/21147/2010

A research thesis submitted in fulfillment of the requirements for the award of Doctor of Philosophy Degree (Microbiology) in the School of Pure and Applied Sciences of Kenyatta University

June, 2016

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university or for any other award

Mwajuma June Jeophita
Department of Microbiology
Kenyatta University

Signature _____ Date 6/6/16

We confirm that the work reported in this thesis was carried out by the candidate under our supervision

SUPERVISORS:

Prof. Paul Okemo
Department of Microbiology
Kenyatta University
P.O. Box 43844 - 00100, Nairobi

Prof. Kiplagat Kotut
Biological Sciences Department
Embu University College
P.O. Box 6 - 60100, Embu

Signature _____ Date 13/6/16 Signature _____ Date 7/6/16

Dr. Matoka Charles Mboya
Dept of Pure and Applied Sciences
Technical University of Mombasa
P.O. Box 90420 - 80100, Mombasa

Dr. Carren Moraa Bosire
Dept of Pure and Applied Sciences
Technical University of Mombasa
P.O. Box 90420 - 80100, Mombasa

Signature _____ Date 3/6/16 Signature _____ Date 7/6/16

DEDICATION

To my children Neema, Baraka, Wema and Ahsante; "If heaven is missing four angels I know exactly where they are."

ACKNOWLEDGEMENTS

I wish to express my gratitude to everyone involved in this process. I thank my supervisors Prof. Paul Okemo, Prof. Kiplagat Kotut, Dr Charles Mboya and Dr Carren Bosire for providing me excellent guidance, constructive critism and support. To the late Dr Alexander Njue, I am thankful to for introducing me to the world of water and what an interesting world it has turned out to be. I wish to acknowledge Dr Maingi and Dr Kebira for all the help with administrative issues during my studies and finalising this thesis.

I thank the lab technologists Nobert, Joan and Ken for creating an inspiring working environment from the wilderness of Baolala to the minispins of L7. In a very special way I would like to thank my dream team Furaha, Dr. Najya and Mangeni for sharing the wilderness and with an exceptionally high-quality sense of humour. Iam greatly indebted to Dr. Scot of M.R. DNA, Shallowater, TX for provision of the sequencing and genomics platform.

Dr.s Annekelly, Odilia and Amanuel are specifically acknowledged for their patient and long term contribution in introducing me to sequencing, Linux, and bioinformatics. Those Ubuntu and QIIME moments are what have shaped the scientist in me. I acknowledge the financial support accorded to me by NACOSTI in carrying out this study and the logistical support of the Coast Water Services Board with its system engineers.

My husband Steve is thanked for efficiently disconnecting me from my thesis with the seafront 'chomas' in recent months - you're Godsend. I am grateful to Pastors Kipruto

and Maria, for praying with me and celebrating me every step of the way. My friends 'Ofisi', thanks for being so exceptionally lovely and always there for me, your unscientific input was invaluable. Glory to God for making everything beautiful in its own perfect time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ANOSIM	Analysis of similarities
ATP	Adenosine Triphosphate
CCA	Cannonical Correspondence analysis
CD	Candidate division
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic Acid
DWDS	Drinking water distribution systems
EDTA	Ethylene Diamine Tetra-Acetic Acid
NMDS	Non metric dimensional scaling
NPMANOVA	Non-Parametric multivariate analysis of variance
OTU	Operational taxonomic units
PCA	Principal Componenents analysis
PCoA	Principal Co-ordinates analysis
PCR	Polymerase Chain Reaction
R2A	Reasoner´s 2 Agar
RDA	Redudancy analysis
RDP	Ribosomal Database Project
RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulfate
UB	Uncultured bacteria

ABSTRACT

Hygienic problems in drinking water distribution systems may originate from contamination by external microorganisms or growth of indigenous biomass. Among all water related disease outbreaks, 33% waterborne diseases were caused by contaminated source water, 39% by inadequate or interrupted treatment processes and 18% by distribution systems and premise plumbing deficiencies (Craun *et al.*, 2010). The Mombasa County water distribution systems have since inception undergone a sanitary survey and there has been no information on the bacterial community composition therein. The primary focus of this study therefore was to generate detailed information on the genome information stored within the water distribution system microbiome. This was achieved through an analysis of the 16S rRNA phylotypes of the microbes present within the microbial community of 21 water and biofilm samples collected from Mzima and Baricho water lines, which have been operational for more than 50 years, making them excellent sources of mature distribution system biofilms. Physico-chemical characteristics of the water were determined by a portable Palin Photometer and Atomic Absorption Spectrophotometer; while bacterial taxonomic affiliations were analyzed using 16S rRNA based 454-FLX Titanium pyrosequencing. Water from both water lines registered pH, nitrates, phosphates and residual chlorine levels that were within the limits stipulated for drinking water by the Kenya Bureau of Standards (KEBS). All Baricho line samples had iron and lead levels way above the maximum allowable limit of 0.05 mg L^{-1}. Nitrates, iron and temperature correlated positively with bacterial community composition and diversity in all samples as shown by Canonical correspondence analysis (Mantel test: $r = 0.27$, $P = 0.001$). Pyrosequencing yielded 27,937 sequences, which were denovo clustered into 2,294 unique operational taxonomic units (OTUs) based on their sequence similarity (3%). A total of 20 bacterial phyla and 6 candidate phyla were identified from pooled samples and were dominated by the *Proteobacteria* (73.2%), *Firmicutes* (13.4%), *Bacteriodetes* (5.9%) and Candidate divisions at 0.9% of thetotal phylotypes. Rényi diversity profiles demonstrated that all sampled sites regardless of source or type have larger species richness, but lower species evenness. Shannon–Wiener diversity (*H'*) index of each sampling site ranged from 0.00 to 1.725. The highest bacterial diversity was found in Baricho water (1.725) and Mzima biofilms (1.391). Biofilms featured characteristically higher bacterial diversity, richness and abundance than bulk water. Bulk water was predominated by *Nitrospirae* (20.2%), *Betaproteobacteria* (15.9%) and *Alphaproteobacteria* (6.4%), while *Favobacteria* (8.6%), *Deltaproteobcteria* (3.2%), bacteria *NPL.UPA2* and Candidate division *ODI* were characteristic of biofilms. Redundancy analysis indicated substantial comparative differences in water and biofilms bacterial community composition among the two water lines. *Nitrospirae, Elusimicrobia, Cyanobacteria, Gemmatimonadetes, NP-UPA2,* and Candidate Divisions *TM7, OP11* and *OD1* were present only at source but not at endpoints. Differences in community structure and abundance were noted between Baricho water line source and endpoint water and biofilms bacterial composition ($p= 0.013$). A total of 140 phylotypes of potentially pathogenic species including; *Pseudomonas, Escherichia, Shigella, Aeromonas, Enterobacter* and *Bdelovibrio* were identified. Metagenome analyses also confirmed the ubiquity of mycobacteria in drinking water distribution

systems. Maintainance of the integrity of water systems, periodic monitoring and effective treatment techniques should take precedence in water delivery services to reduce the risk of contaminating the drinking water with pathogenic microorganisms.

CHAPTER 1: INTRODUCTION

1.1 Background

Microbial contamination of water resources continues to be a major threat to human health. The presence of microorganisms in drinking water distribution systems causes problems of corrosion, quality deterioration, and outbreak of waterborne diseases, of which public health risks associated with the outbreak of waterborne diseases is of the most concern (Craun *et al.*, 2010; Li *et al.*, 2010). About 88% of diarrheal disease, which are estimated to account for 4.1% of the total daily global burden of disease and the annual deaths of 1.8 million people, (mostly children in developing countries), are attributable to the consumption of water with microbial contamination (WHO, 2009). Surveys conducted by the Center for Disease Control and Prevention (CDC) show that from the year 1971 to 2006, 780 disease outbreaks were associated with drinking water contamination. These resulted in 577,094 cases of illness and 93 deaths. Among all the outbreaks, 33% waterborne diseases were caused by contaminated source water, 39% by inadequate or interrupted treatment processes and 18% by distribution systems and premise plumbing deficiencies (Craun *et al.*, 2010). It is therefore very important to understand the survival, growth and proliferation of microorganisms in drinking water and water distribution systems for the surveillance, assessment, control, and prevention of public health risks associated with drinking water supply.

Microbial contamination challenges associated with drinking water distribution systems include biofilm growth (Camper, 2004), nitrification (Regan *et al.,* 2003), microbe mediated corrosion (Beech and Sunner, 2004), and the persistence of pathogens

(Emtiazi *et al.,* 2004). Microorganisms are always present in drinking water no matter how rigorous the precautions taken during the production and distribution. Drinking water transported through the distribution networks is subject to both chemical and microbial quality changes. A variety of microbes, including pathogens, are able to proliferate in such an environment, forming biofilms on the interior surface of the pipes. This occurs despite the presence of disinfectant residues, although drinking water distribution systems are generally perceived as unsuitable for microbial growth since they are oligotrophic. In biofilms, microbes are embedded in a matrix of extracellular polymeric substances, which protect them from disinfectants and high-shear flow. They also provide nutrients for microbes by accumulating organic and inorganic debris from external sources through adsorption (Brown *et al.,* 1995; Hoyle *et al.*, 1992). In water distribution systems, biofilms periodically break off and cause hygienic problems of drinking water (Stoodley *et al.,* 2002). In the presence of biofilms, drinking water distribution systems could encounter operational problems such as pipe corrosion, water quality deterioration brought about by contamination by undesirable substances. In addition, hygienic problems may also originate from contamination by external microorganisms or growth of indigenous biomass (Hu *et al.,* 2005).

Identifying the members of a core microbiome (the suite of members shared among microbial consortia from similar habitats) is critical to understanding the stability and components consistency across complex microbial assemblages (Shade and Handelsman, 2012). The hunt for a core microbiome of the water distribution system correlates with the patterns of beta-diversity (differences in community composition

among sites) and distribution of microorganisms across separated sites of the system. To unravel this great diversity, modern and more powerful estimation methods, such as high through put molecular extraction and next generation sequencing (Rothberg *et al.,* 2011; Andersen *et al.,* 2010; Bentley *et al.,* 2008; Bentley, 2006; Margulies *et al.,* 2005) have been applied.

Water distribution systems biofilms have received much focus due to their importance in potable water delivery to end-point consumers (Feazel *et al*., 2009). Both culture-based and culture independent molecular approaches have been used to establish the bacterial communities of different locations or pipe materials within a water distribution system and diverse sources and nutrient levels. Studies on the bacterial communities of a water distribution system have indicated that populations can differ significantly from source water to tap water (Eichler *et al*., 2006). Activities of microbial growth and the presence of potential opportunistic pathogens have been detected in tap water, faucets, and showerheads, which are end points of the water distribution systems (Kormas *et al*., 2010; Feazel *et al*., 2009). Many studies attribute the survival of pathogens to resistance mechanisms like cell wall permeability and biofilm formation (Freeman *et al*., 2006; Grobe *et al*., 2001), for instance the type of bacterial communities present in a water distribution systems and the disinfection regime applied may also influence pathogen retention in the water distribution systems. Nitrifying microorganisms, for instance, can contribute to the depletion of monochloramine and subsequently lead to increased overall microbial growth (Eichler *et al*., 2006).

Compared to the analysis of microbial communities in source water and end-point drinking water, investigations related to the delivery process are difficult due to limited access and the high cost involved in sampling within the water distribution systems. Earlier studies have reported that episodic chlorination may accelerate the development of microbial communities with increased resistance to disinfectants (Codony *et al.*, 2005). Studies als show that bacterial diversity can also affect disinfection efficacy and pathogen survival (Batte *et al.*, 2003). Many studies have used surrogates such as model distribution systems and removable coupons for biofilm attachment inserted for short times, in real distribution systems. Studies suggest that biofilm development may require several years before a steady state is achieved, which limits the relevance of short-term model studies (Martiny *et al.,* 2003). The occurrence of bacterial community succession was observed in a model water distribution systems (Martiny *et al.*, 2003), which indicated the importance of long-term monitoring and suveillance of bacterial biofilm development in the water distribution systems. These findings also demonstrated the need for the direct evaluation of microbial communities within the water distribution systems, which would complement studies of the source water and end points. Free chlorine is used as the primary disinfectant in most drinking water distribution systems. However, chlorine disinfection promotes the formation of disinfectant by-products (DBPs). Chlorination aims at mitigating the presence of pathogens, but does not completely eradicate the growth of microorganisms in water distribution systems. Diverse microbial communities have been shown to inhabit water distribution systems.

As drinking water systems distribute treated water, they should be devoid of microorganisms, especially waterborne pathogens. This is because the goal of water treatment is to minimize microorganisms, besides eliminating waterborne pathogens. Early studies relied upon relatively small sample sizes to estimate microbial diversities in the population (Colwell and Coddington, 1994; Schmidt *et al.*, 1991; Trudel *et al.*, 1986). The implication of this is that in the event of contamination, conventional methods used for routine microbial water quality testing may not detect the same. Only high-throughput and high-resolution detection methods, such as those offered by next-generation sequencing (NGS) techniques, can adequately assist in the task of extensively and intensively investigating patterns of distribution of microbial communities in this environment.

Although next generation sequencing technologies are not exempt from a number of potential sources of bias (Gihring *et al.*, 2012) they allow the investigation of microbial diversity at an unprecedented level of resolution. By contrast, while allowing one to gain information on how the dominant members of microbial communities differ in composition across ecological landscapes. Diversity profiling techniques suffer from inherently low levels of taxonomic discrimination (Griffiths *et al.* 2011). Deep sequencing allows an overview of the communities as a whole, overcoming the limited views experienced using previously employed techniques.

The adaptation of culture-based assays (Figueras and Borrego, 2010) and culture-independent approaches, such as 16S rRNA that target gene sequence analysis using Sanger chemistry (Revetta *et al.*, 2011) and pyrosequencing (Hong *et al.*, 2010) have

been use to show the diverse microbial communities inhabiting water distribution systems. Additionally, fluorescence *in situ* hybridization (FISH) targeting the 16S rRNA gene has been used to detect active bacteria in drinking water biofilms (Williams and Braun-Howland, 2003). Most of these previously used methods are generally limited in scope by targeting specific indicator species. For example, culture-based techniques are biased toward a small fraction of the inhabiting microbiota. On the other hand, most studies using DNA-based approaches have targeted phylogenetic genes, which provide limited information on the public health relevance of the microbial groups detected.

The distribution system is the remaining component of public water supplies yet to be adequately addressed in national efforts to eradicate waterborne disease. The ecology of the distribution system is poorly understood, making risk assessment via pathogen occurrence measurements difficult. There is very little information available about the types, activities, and distribution of microorganisms in distribution systems. Limited heterotrophic plate count data are available for some systems, but these data are not routinely collected, they underestimate the numbers of organisms present, and they include many organisms that do not necessarily present a health risk. Recently completed epidemiological studies have either not focused on the specific contribution of distribution system contamination to gastrointestinal illness, or they have not identified a direct link between illness and sources of drinking water.

It is against this background that the study seeks to adopt ultra modern molecular biology techniques for evaluation of bacteria composition and diversity in the drinking

water distribution systems bulk water and pipe interface biofilms and their possible link to waterbourne disease, so as to guide in the mitigation of the associated health risks.

1.2 Problem Statement

Recent survey on disease outbreaks have revealed that a significant portion were associated with drinking water contamination. Among all the outbreaks, 33% waterborne diseases were caused by contaminated source water, 39% by inadequate or interrupted treatment processes and 18% by distribution systems and premise plumbing deficiencies (Craun *et al.*, 2010). Cases of cholera and other water associated diarrhoeal diseases continue to cause death and suffering in Mombasa County, raising a lot of querries on the quality of water consumed by the public. Whereas a lot of work has been done on the contamination and microbial composition of source water using the traditional culture methods, information on water distribution systems remains scanty. The traditional culture dependent methods of microbe isolation and identification ignore substantialy viable nonculturable populations.

The analysis of microbiological biodiversity has advanced significantly with the inclusion of Next Generation Sequencing (NGS) technologies (Shokralla *et al.*, 2012). Modern sequencing technology allows data retrieval from possibly thousands or even millions of microorganisms, hence rendering more reliable and comprehensive studies (Edwards *et al.*, 2006). This capability can be further exploited by combining several samples in a single experiment using tag-encoded amplicon pyrosequencing (Binladen *et al.*, 2007) which permits subsequent binning of data to samples. Traditionally, bacterial biodiversity has been measured by estimating the number of culturable species

present in the population. However, it is now possible to collect genetic information from unculturable organisms, which frequently cannot be ascribed to characterized species. Operational Taxonomic Units (OTUs) at 3%, 5% and 10% dissimilarity are commonly used to estimate richness of species, genus and phylum respectively (Creer *et al.*, 2010).

The purpose of this study was to characterize the bacterial communities of operational drinking water distribution networks. The study also sought to determine the diversity of bacteria and their role in microbiologically induced water quality change and the success of water treatment. Use of molecular microbial ecology tools was employed to improve the understanding of the total microbiota, since the conventional culture-based methods are selective and do not cover the entire microbial diversity of complex environments. Pyrosequencing provides rapid and accurate screening of total bacterial populations, and the opportunity to sample biofilms from a real distribution system, provided needed insights on bacterial community diversity in these engineered systems.

1.3 Justification

The fundamental goal in microbial ecology is to characterize all microbial communities, and to explain how abiotic and biotic environmental factors and interactions impact these communities (Gentry *et al.*, 2006). Information on the water distribution systems microbiome composition will enhance the understanding on the core populations, occurrence of pathogenic bacteria, and their growth in the presence of disinfectants. This will serve as guide in designing innovative and effective control strategies that will ensure safe and high quality drinking water (Xi *et al.*, 2009).

The scale of the problems caused by the activity of the indigenous microorganisms remains virtually unknown, since the monitoring frequency by the water supplies in the distribution networks is generally sparse and the microbial techniques used are inadequate for the characterization of the indigenous microorganisms. Several novel bacterial strains have been isolated from municipal distribution systems during the process of characterizing microbial diversity (Rickard *et al.,* 2005; Kalmbach *et. al.*, 1997). A rigorous characterization of these strains has been achieved in this study.

Local water quality surveillance employs culture techniques. It has been shown that less than 1% of environmental bacteria are able to grow in the laboratory (Hugenholtz *et al.*, 1998). This would leave the remaining 99 % of environmental bacteria out of reach if culture independent methods like metagenomics did not exist. Application of Next-generation sequencing to this study provided an enabling technology that allows for the study of the genetic makeup of such microbial communities at a higher taxonomic

resolution that can enable interpretation of both biogeographic patterns and the processes driving such patterns.

Waterborne pathogens that are able to persist and reproduce in the distribution system can cause infections of the gastrointestinal tract, skin and lymph nodes (Szewzyk *et al.,* 2000). The distribution system pathogens comprising the *Mycobacterium avium* complex, for example, have been identified as the most common source of bacterial infection in AIDS patients (Kunimoto *et al.,* 2003). Microbes in biofilms, rather than those in the bulk water, are mostly responsible for the deterioration of the drinking water quality during the drinking water transport (Block, 1992). It is therefore critical to know the nature of biofilms formed in water distribution systems for the protection of public health.

The appearance of potential antibiotic resistance in drinking water distribution systems requires increased surveillance for risk assessment and prevention strategies to protect the health of the public. With metagenomic and computational techniques, it is possible to identify antibiotic resistance bacteria and their genes in environmental microbiomes. The availability of metagenomic data sets not only enables the study of known antibiotic resistance genes in such environments, but also provides the possibility of finding new, previously undiscovered antibiotic resistance genes. Members of rare biospheres such as Water distribution systems, may act to form a seedbank that provide the genetic potential to respond to dramatic ecosystem changes, and thus any rare species may become a dominant species in the future.

Using the metagenome-based approaches, this study offers a more comprehensive view of the genetic complexity of natural and engineered microbial communities. This allows for better assessment of the bacterial taxonomic diversity and metabolic potential within the distribution system community.

1.4 Null Hypotheses

i) The bacterial composition and diversity of the drinking water distribution systems of Mombasa is not different from the typical bacterial flora of water distribution systems

ii) There is no difference in the relative abundance of different bacterial community members in Baricho and Mzima water distribution systems.

iii) There is no relationship between the bacterial community composition of drinking water distribution systems and its physico chemical quality.

iv) The bacterial communities of biofilm and bulk water populations are similar.

v) There is no variation in bacterial composition and diversity at the source and end point phases of the different of the water distribution systems.

1.5 Objectives

1.5.1 The General Objective

To determine the community structure of bacteria in Baricho and Mzima drinking water distribution systems in Mombasa County using metagenomic approaches.

1.5.2 Specific Objectives

i) To determine the bacterial community composition and diversity in water distribution systems of Mombasa.

ii) To estimate the relative abundance of different bacterial taxa in Baricho and Mzima water distribution systems.

iii) To establish the relationship between bacterial composition and the physico chemical characteristics of water.

iv) To compare the bacterial composition and abundance in biofilm and bulk water.

v) To assess the changes in bacterial community composition at different phases of the water distribution systems.

CHAPTER 2: LITERATURE REVIEW

2.1 Microbial Contamination of Water Distribution Systems

Water distribution systems are generally perceived to be unsuitable for microbial growth because of their oligotrophic and high-shear conditions, plus the presence of disinfectant residues. However, various microbes are able to survive in this environment, forming biofilms on the pipe interior surface and embedding in a matrix of extrapolymeric substances, which protects them from disinfectants and high-shear flow (Brown *et al.,* 1995; Hoyle *et al.,* 1992). The conventional approach to water treatment, which involves maintaining a disinfectant residual, is often ineffective at controlling microbial growth (LeChevallier *et al.*, 1996).

The biological and physico-chemical processes that take place during drinking water distribution, at the consumers' taps generally results in a lower quality than the treated water at the treatment plant (Verberk, *et al.,* 2007; Vreeburg and Boxall 2007; Lee, O'Conner *et al.,* 1980). Most challenges in water distribution systems such as pipe wall biofilm growth (Camper 2004), nitrification (Regan *et al.,* 2003 and 2002), bio-corrosion of pipe material (Beech and Sunner, 2004; Lee O'Conner *et al.,* 1980), deterioration of taste and odor (Hoehn 1988) and proliferation of opportunistic pathogenic bacteria are microbial in nature, (Feazel *et al.,* 2009; Emtiazi *et al.*, 2004).

Microbes may grow on the submerged surface in drinking water distribution systems, forming a biofilm, which may provide favorable conditions for the growth of pathogens (Parsek and Singh, 2003). These organisms can detach from the pipe surface with biofilms threatening the safety of drinking water (Obst *et al.*, 2006; Flemming *et al.,*

2002; Stoodley *et al.,* 2002). Many factors have been reported to account for biofilm growth in drinking water distribution system (DWDS), including water source, environmental and hydraulic conditions (Lehtola *et al.,* 2004; Angles *et al.,* 1999).

It has been estimated that over 90% of microorganisms in nature are present in the form of biofilm (Flemming *et al.,* 2002; Costerton *et al.,* 1995). Previously, biofilm was considered to be a homogeneous layer covering the substratum. However, the introduction of confocal scanning laser microscopy (CSLM), which enables study of biofilm in its natural hydrated state and allows for observation of thin optical sections across the biofilm profile generating three-dimensional image of the biofilm, has changed this perception (Wimpenny and Colasanti, 1997). The heterogeneity of the biofilm permits formation of different microenvironments and thereby different ecological niches inhabited by specialized microorganisms.

2.2 Complexities Associated with Controlling Microbial Growth

Optimizing the management of distribution systems and controlling microbial growth is difficult because of the complexity of these systems. The survival of microorganisms is based upon the interactions of many variables, including temperature (Ndiongue *et al.,* 2005), pipe surface (Lehtola *et al.,* 2005), nutrient levels (Wijeyekoon *et al.,* 2004; Butterfield *et al.,* 2002) and type and concentration of disinfectants (Norton *et al.,* 2004).

A number of studies have confirmed the complex behavior of micro-orgnanisms in drinking water. Flow chamber studies have concluded that the presence of high

concentrations of disinfectants is not sufficient to eliminate the survival of pathogens, including *Legionella pneumophila* and *E. coli* (Williams and Braun-Howland, 2003). In a related study, Langmark *et al.* (2005) found out that the application of two common disinfectants, monochloramine and ultra-violet light did not deter *L. pneumophila* from accumulating in biofilms in a pilot-scale distribution system. In a study focusing on accumulated growth, Pozos *et al.* (2004) observed that biofilms exposed to strains of *E. coli* and *Klebsiella pneumoniae* developed stable populations of both opportunistic pathogens proportional to the biofilm density of heterotrophic bacteria. Another study focusing on the behaviour of *Mycobacterium xenopi* in biofilms revealed that it exhibited long-term persistence in biofilms and that a steady concentration of *M. xenopi* cells was returned to the water column from biofilms (Dailloux *et al.,* 2003).

2.3 Biofilm Biology

Biofilm formation is a dynamic process and comprises a serries of steps. The process starts with preconditioning of the adhesion surface. The preconditioning film is a thin layer of organic molecules and ions covering the adhesion surface that is formed before any microorganisms attach to the surface. These molecules may adhere to the surfaces through physical or chemical adsorption. The strength of biofilm adhesion is largely dependent on the cohesion of this conditioning (Bos *et al.,* 1999; Busscher *et al.,* 1995). Effective adhesion of microorganisms to surfaces starts by the transport of microbial cells to the adhesion surface either by fluid dynamics, gravitational forces and Brownian motion, or by migration through active cell motility.

The attached microorganisms start growing and form microcolonies, excreting organic polymers and initiating the formation of the biofilm matrix. Exopolysaccharide synthesis has been shown to be very important in the formation of microcolonies (Watnick and Kolter, 1999). As biofilm thickness increases, transport of nutrients from the external liquid media to the inner layers of biofilm and transport of excreted metabolites in the opposite direction are important for biofilm maintenance. This leads to the development of a complex and organized consortia of microorganisms embedded in an organic matrix that protects the microorganisms inside from stress factors. It is at this stage that microorganisms produce large amounts of EPS. The structure of a mature biofilm depends on the microbial composition, EPS production, the nutrient availability, hydrodynamic conditions and temperature. In a mature biofilm, several processes may occur simultaneously: bacteria detachment into water, attachment of planktonic bacteria, growth and death. However, in this stage these processes are at equilibrium and the number of attached cells per unit surface area is constant in time, although with periodic fluctuations (Bryers, 2000; Vieira *et al.,* 1993).

Hydraulic shear stress provoked by high flow velocities can lead to detachment of bacteria and biofilm aggregates (sloughing), with higher detachment rates at increasing shear (Characklis *et al.,* 1990). Detachment occurs due to different mechanisms: erosion (the continuous release of single cells or small clusters of cells), sloughing (the rapid detachment of large portions of the biofilm), abrasion (collision of solid particles with the biofilm), and predator grazing. Erosion and sloughing can result from biofilm-associated processes such as enzyme production (Allison *et al.,* 1998), excretion of

certain signalling molecules (Wuertz *et al.,* 2004), cell-cycle mediated events (Gilbert *et al.,* 1993; Allison *et al.,* 1990), and the excretion of surface modified products (surfactants) by certain bacteria (Rosenberg and Ron, 1999). It may also arise from external factors such as shear forces (Picioreanu *et al.,* 2001), variations in the nutrient concentration (Sauer *et al.,* 2004), chemical change in EPS due to the presence of chelating agents.

The biofilm matrix (composed by all inorganic and organic substances surrounding the cells) has several functions. Acting as a structural backbone, the biofilm matrix protects bacteria from being washed out, from mechanical shocks, from toxic/lethal attacks by antibiotics (Stewart and Costerton, 2001), disinfection chemicals (Costerton *et al.,* 1995; LeChevallier *et al.,* 1988a), UV radiation (Hijnen *et al.,* 2006), predators (Keevil and Walker, 1992) and from desiccation (Sutherland, 2001; LeChevallier *et al.,* 1988b). The matrix promotes the storage of nutrients for intake during periods of limitation (Wolfaardt *et al.,* 1998), retention of extracellular enzymes (Flemming, 2002), horizontal gene transfer (Molin and Tolker-Nielsen, 2003), and exchange of signalling molecules and metabolites (Wuertz *et al.,* 2004).

2.4 The Importance of Biofilms

Biofilms are suspected to be the primary source of microorganisms in distribution systems that are fed by adequately treated water and have no pipeline breaches (LeChevallier *et al.,* 1987) and are a particular concern in older distribution systems (Geldreich, 1996). In a recent study of distribution systems in Parisian suburbs, it was found that biofilms attached to the surface of a 100 mm diameter pipe contained 25

times more bacterial cells per unit length than the adjacent bulk water (Servais *et al.*, 2004). Biofilms predominate because attached cells have certain advantages over planktonic cells, such as the ability to metabolize recalcitrant organic compounds (Camper, 2004) and increased resistance to chlorine and other biocides (Tachikawa *et. al.*, 2005; Emtiazi *et al.,* 2004).

When pathogenic bacteria enter a distribution network, they are retained in biofilms where they can remain for a prolonged period of time. The biofilm may act as a safe haven for the pathogenic bacteria since they are able to survive or even grow in it. In addition, the biofilm bacteria have been shown to be more resistant to disinfection (Armon *et al.,* 1998) hence protecting these harmful organisms against disinfectants like chlorine. Some pathogens have a potential for growth in distribution networks. In a study by Szewzyk *et al.* (1994) pathogenic *Escherichia coli* was able to grow in a benzoic acid degrading biofilm and re-enter the bulk water. Recent studies have shown that *Helicobacter* sp. can be present in seemingly uncontaminated drinking water biofilms (Park *et al.,* 2001).

2.5 Antibiotic Resistance Genes in Drinking Water

Modern health care is highly reliant on effective antibiotics to treat bacterial infections. However, bacteria are becoming resistant to many of the commonly used antibiotics, hence the need for careful usage of antibiotics (Andersson and Hughes, 2012 and 2010,). Microbial organisms have a long history of withstanding antibiotics (D'Costa *et al.*, 2011; Davies and Davies, 2010). Bacteria have several mechanisms with which they can exchange genetic material with other bacteria through a process called horizontal

gene transfer (Bennett, 2008; Aminov and Mackie, 2007). This enables bacteria to share genes for antibiotic resistance. Many common bacterial antibiotic resistance genes originate from the environment (Allen *et al.*, 2010; Cantón, 2009). There is therefore increasing resistance rates and the potential impact of antibiotic residues in aquatic environments in the recent past (Kemper, 2008; Wright, 2007; Sarmah *et al.*, 2006). Apart from the chemical pollution caused by antibiotics themselves, use of antibiotics may also accelerate the development of antibiotic resistance genes (ARGs) in bacteria, which could pose health risks to humans and animals (Kemper, 2008). There is growing evidence that clinical resistance is intimately associated with environmental ARGs and bacteria (Abriouel *et al.*, 2008; Prabhu *et al.*, 2007).

Prevalence and resistance patterns of various microbial genera isolated from drinking water distribution system have been recently reported (Koksal *et al.*, 2007; Ram *et al.*, 2008). Multiple-antibiotic-resistant *E. coli* strains isolated from drinking water was found to carry ARGs encoding resistances to aminoglycoside, β-lactam, tetracycline, and trimethoprim-sulfamethoxazole (Alpay-Karaoglu *et al.*, 2007; Cernat *et al.*, 2007), as well as class 1 integrons (Ozgumus *et al.*, 2007). In order to indicate possible ARGs transfer from wastewater and surface water to the drinking water distribution network, Schwartz *et al.*, (2003) and Obst *et al.*, (2006) investigated biofilms in hospital and municipal wastewater, as well as drinking water from river bank filtrate, and found that vanA and ampC genes occurred not only in wastewater biofilms but also in drinking water biofilms. Florfenicol resistance gene floR and penicillin resistance gene penA has also been found in *Listeria monocytogenes* isolated from drinking water in dairy farms

(Srinivasan *et al.*, 2005). These bacteria might be transmitted from environment to human via direct or indirect contact (Rodríguez *et al.*, 2006; Iversen *et al.*, 2004).

2.6 Microbial Community Diversity in Water Distribution Systems

A recent analysis of the bulk water of a chlorinated DS found that Gram-positive bacteria and Alpha-, Beta- and Gammaproteobacteria constituted the major groups among heterotrophic isolates (Tokajian *et al.,* 2005). Alphaproteobacteria are the dominant isolates in both chloraminated and chlorinated water from model Distribution systems, whereas Betaproteobacteria have been found to be more abundant in chloraminated water than in chlorinated water (Williams *et al.,* 2004). 16S rRNA gene-directed PCR and denaturing gradient gel electrophoresis (DGGE) revealed that Betaproteobacteria were also abundant in biofilms of non-chlorinated distribution systems (Emtiazi *et. al.*, 2004).

A variety of microbes found in biofilm are nonpathogenic bacteria according to DNA sequences analysis of HPC isolates and clone libraries (Martiny *et al.,*2005; Lee *et al.,* 2005; Williams *et al.,* 2005 and 2004; Santo Domingo *et al.,* 2003; Lee and Kim, 2002). *Hyphomicrobium spp.* has been widely found in chloraminated and chlorinated water (Williams *et al.,* 2004), and mineral drinking water. However, no pathogenic species have been isolated from this group so far. Although non-pathogens do not cause diseases directly, they are also important for the drinking water safety since pathogens can find favorable conditions for their proliferation in the biofilm formed by the non-pathogens (Parsek and Singh, 2003; Rogers *et al.,* 1994; LeChevallier *et al.,* 1987).

Pathogens, such as *Legionella bozemanii*, *Mycobacterium smegmatus*, and *Cryptosporidium parvum* readily colonize the biofilm formed by *Acidovorax spp.* and *Bacillus spp.* (Arrage and White, 1997). In water distribution systems some copiotrophic bacteria may also survive in biofilms even though the bulk water is oligotrophic. This is because they can use remnants and excretions of oligotrophic bacteria as sources of carbon and nutrient. For example, the presence of oligotrophic bacteria was reported to enhance the survival of *Aeromonas spp.* in water distribution systems biofilms (Messi *et al.*, 2002).

Studies of model distribution systems by 16S rRNA gene sequence analysis show that biofilm species richness was comparable to the species richness in the bulk water during the initial stages then decreased as a dominant bacterium related to *Nitrospira* colonized the surfaces, comprising 78% of the biofilm cells. Biofilm species richness increased again as a stable biofilm community composition was achieved after almost two years (Martiny *et al.*, 2003). Consistent with this observation, other studies with model Distribution systems suggest that as biofilms age, cell density stabilizes and species diversity increases (Lee *et al.*, 2005).

2.7 Factors Affecting Biofilm Growth in Water Distribution Systems

Biofilm growth in water distribution systems is caused by the microbial utilization of biodegradable organics and nutrients, mainly nitrogen and phosphorus (Van der Kooij *et al.*, 2005; Bachmann and Edyvean, 2005; Charnock and Kjønnø, 2000). The estimated proper molar ratio of carbon, nitrogen and phosphorus required for bacterial

growth varies from 100:10:1 to 100:25:4 (Bachmann and Edyvean, 2005; Chandy and Angles, 2001; Ollos *et al.,* 1998; USEPA, 1992). Since both N and P are relatively abundant in most treated drinking water, organic carbon is often considered as the limiting factor (Bachmann and Edyvean, 2005).

Nitrate nitrogen levels in drinking water normally ranges from 0 to 10 mg L^{-1}, with an average of about 1.4 mg L^{-1}, a level which is sufficient for biofilms growth in water distribution systems (Donlan, 2002; Miettinen *et al.,* 1997). Phosphate concentration in water distribution systems is normally below 1 mg L^{-1} (Mannuel *et al.,* 2007; Lehtola *et. al.*, 2006). Adding phosphate may affect the biofilm growth and the biomass content in bulk water in water distribution systems, depending on the Asimilable organic carbon (AOC) and biodegradable dissolved organic carbon (BDOC) concentrations (Park *et al.,* 2008; Volk and LeChevallier, 2000).

Pipe materials in water distribution systems can generally be classified into three types; cement (either solid or coated), metals and plastics. Niquette *et al,* (2000) reported that metals (tarred steel and grey iron), produced the most amount of biofilm followed by cements including cemented steel, cemented cast iron and asbestos cement. Plastics such as polyvinyl chloride and polyethylene produce the least amounts of biofilm. Corrosion products, such as iron oxides, might adsorb and concentrate organic carbon and nutrients in the water, such as humic substances, to promote biofilm formation (Camper, 2004; Butterfield *et al.,* 2002; Gu *et al.,* 1994). The scale (rust) of corrosion could provide more surfaces for biofilm growth and reduce shearing force to avoid biofilm detachment. Some bacteria can also directly utilize ferrous iron or hydrogen

produced from iron corrosion as energy source for growth (Videla and Characklis, 1992). PVC and cement pipes are generally preferred over metal pipes to avoid corrosion and reduce bacterial growth in water distribution systems (Niquette *et al.*, 2000).

Disinfection by chlorination, UV irradiation or ozonation is the traditional method for microbial control in water distribution systems (Biswas *et al.*, 1993). Chlorination by either free chlorine or chloramines is the most common disinfection process. Studies have shown that biofilm densities decrease exponentially with the increase of residual free chlorine (Ndiongue *et al.*, 2005). Application of 0.57 mg Cl_2 L^{-1} to chlorine-free water was reported to decrease bacteria in biofilm from 5.8 to 0.3 CFU cm^{-2} (Tsai, 2006). Increasing free chlorine from 0 to 0.5 mg Cl_2 L^{-1} reduced biofilm HPC levels by about four orders of magnitude (Ollos *et al.*, 2003). Disinfection with chlorine dioxide and chlorite, can reduce the concentration of planktonic bacteria, but has little to no effect on the concentration of biofilm bacteria (Gagnon *et al.*, 2005). The mechanism behind the observed resistance of biofilm cells to disinfection is not known, although hypotheses include mass transfer resistance (Stewart *et al.*, 1996), the formation of persistor cells (Roberts and Stewart, 2005), and protection owing to the production of extracellular polymeric substances (Allesen-Holm *et al.*, 2006).

Temperature plays important role in biofilm growth in water distribution systems (Lund *et al.*, 2003; Smith *et al.*, 2000; Ormerod, 1995; Havelaar *et al.*, 1990). Results from coliform occurrence tests in drinking water over a wide range of 0-48°C indicated that systems under 10 – 20 °C had the most coliform, about 2.5 – 3.0 fold higher than that of

under 0 – 5 °C (LeChevallier *et al.,* 1996). The amount of biofilm formed under a temperature range of 36 – 38 °C was 2-20 fold of that under 15 – 19 °C based on PLFA analysis (Smith *et al.,* 2000).

2.8 Significance and Public Health Concerns Associated with Biofilms (Pathogens)

The persistence and growth of pathogens is a central concern in distribution systems. Many pathogenic bacteria have been detected in WDS biofilm. These bacteria include, *Pseudomonas aeruginosa* (Lee and Kim, 2002), *Campylobacter sp.* (Buswell *et al.,* 1998), *Acinetobacter sp.* (Nagy and Olson, 1985) and *Escherichia coli* (LeChevallier *et al.,* 1987). These bacteria were identified by the cultivation method while *Legionella pneumophila* (Williams *et al.,* 2004) was identified by use of clone libraries analysis. The presence of *E. coli, Pseudomonas Aeromonas, Artrobacter, Caulobacter, Klebsiella Bacillus, Enterobacter, Citrobacter, Acinetobacter, Prosthescomicrobium, Alcaligenes, Serratior* and *Actinolegionella*has have also been reported. Pathogenic bacteria such as *Pseudomonas, Mycobacterium, Klebsiella, Aeromonas*, *Legionella* spp., *Yersinia enterocolitica*, *Salmonella typhimurium* and enterotoxigenic *E. coli* have also been reported to attach to the pipes surfaces

in water distribution systems (Engel *et al.,* 1980; Wadowsky *et al.,* 1982; Camper *et al.,* 1985). *Mycobacterium xenopi* was found to colonize drinking water biofilms (Dailloux *et al.,* 2003). Other pathogenic bacteria such as *Campylobacter* spp., *Helicobacter pylori* and *Cryptosporidium parvum*in have been reported to have the capacity to survive in biofilms (Armon *et al.,* 1998; Buswell *et al.,* 1998).

2.9 Molecular Analysis of Biofilm Communities

Field surveys using PCR and Southern blot hybridization reported regular detection of pathogens, like *Legionella spp.* and atypical mycobacteria (Emtiazi *et al.,* 2004). *Cryptosporidium spp.* oocytes were detected in bulk water samples and *Helicobacter spp.* were identified in biofilms (Park *et al.,* 2001) in distribution systems using nested PCR methods. Multiplex PCR analysis was used to detect *Mycobacterium avium* and *Mycobacterium intracellulare* as well as several other *Mycobacterium spp.* in water column and biofilm samples (Falkinham *et al.,* 2001). *Aeromonas spp.* have also been found in distribution systems and PCR-based methods have been used to quantify the abundance of specific virulence factor genes in isolated Aeromonas strains in drinking water (Sen and Rodgers, 2004). In addition to the detection of specific pathogens and virulence factors, one study monitored antibiotic resistance genes in distribution system biofilms. Using PCR-based methods, resistance genes responsible for vancomycin-resistance and for βlactamase activities were detected in distribution system biofilms (Schwartz *et al.,* 2003).

The use of molecular tools in the detection of pathogens in drinking water systems, including PCR-based methods, DNA and RNA-targeted hybridizations, and microarray-based technologies, allows for much more sensitive detection of pathogens compared to traditional culture based methods. Research on biofilm has developed and advanced rapidly in recent years, with the application of the denaturing gel gradient electrophoresis (DGGE) and clone libraries. DGGE is usually applied for community comparison (Hoefel *et al.,* 2005; Lee *et al.,* 2005 Emtiazi *et al.,* 2004; Santo Domingo

et al., 2003) while clone libraries are used for bacterial species analysis (Hipsey, 2007; Keinänen-Toivola *et al.,* 2006; Williams *et al.,* 2005 and 2004).

2.10 Molecular Methods of Studying Microbial Diversity

Biological study is technology driven and therefore microbial ecology is not an exception. Major advances in the field have followed revolutionary instrument innovations, such as the microscope, the discovery of DNA, PCR and most recently the advent of next generation sequencing technology (Xu, 2011; Shendure *et al.,* 2005; Margulies *et al.,* 2005). The current range of molecular methods available are extensive, covering all the techniques relying on extracted community DNA, RNA and proteins, but not requiring cultivation of pure microbial cultures (Prosser *et al.,* 2010).

Today, molecular methods are applied to determine the identities and functions of microbes in diverse communities, with a focus on the whole community and its interactions. This is in contrast to traditional methods that typically separated the microbe of interest from its natural physical and biological environment (Prosser *et al.,* 2010). Before the advent of next generation sequencing technologies the term "molecular methods" typically meant either fingerprinting or cloning and sequencing based techniques (Oros-Sichler *et al.,* 2007).

Fingerprinting methods are PCR based techniques that in most cases detect differences in the marker gene among diverse community by electrophoretic separation of produced DNA fragments (Oros-Sichler *et al.,* 2007). The resolution power and identification precision of community members are often relatively constricted. The number of

microbial groups, Operational Taxonomic Units, detected by fingerprinting methods is usually far lower than the actual richness of natural communities with long-tailed rank abundance distribution and the diversity indices calculated based on fingerprints which do not provide reliable information on the true bacterial community diversity (Bent *et al.*, 2007).

These methods therefore work best as a rapid and comparative analysis of multiple samples, for comparing the diversity and community structure of samples with fairly considerable differences along environmental gradients or experimental factors, or for characterizing rather simple and well-studied communities (Nieminen *et al.*, 2012; Juottonen *et al.*, 2008;). With diverse environments, such as seawater and soil with potentially thousands species per gram of sample and potentially billions of microorganisms the resolution power is typically not sufficient (Daniel, 2011; Kirchman, 2008).

2.10.1 DNA Sequencing

DNA sequencing is a process of determining the order of nucleotides in a stretch of DNA. The first sequencing methods were first published in the 1970s when Sanger and colleagues and then Maxam and Gilbert published their sequencing methods (Maxam and Gilbert, 1977; Sanger and Coulson, 1975). Upon its publication in 1977, the Maxam-Gilbert sequencing system became more popular because of the ease of use compared to the first Sanger method. However, later in the same year Sanger and colleagues published a new sequencing strategy relying on sequencing by synthesis and

natural 2'-deoxynucleotides (dNTP) and chain terminating 2', 3'-dideoxynucleotides (ddNTPs) (Sanger *et al.,* 1977). The new Sanger sequencing method soon became the most popular way of determining the nucleotide sequence and later improvements with automation in laboratory work and electrophoresis allowed parallel running of samples and more throughput (Metzker, 2005).

Sanger sequencing method was the first sequencing technique applied in metagenomic studies of natural environments (Hugenholtz and Tyson, 2008). However, despite the automation in methodology, the list of required steps in protocol is long and the stages laborious. This is because the gene of interest needs to be amplified or the environmental DNA fragmented followed by construction of the clone libraries of the resulting amplicons or DNA fragments and then these fragments are sequenced individually (Metzker, 2005). However, Sanger sequencing method has been widely applied with functional genes and community structure and diversity studies using 16S rRNA gene (Rastogi and Sani, 2011), and it is still in use in many laboratories. It is not feasible to sequence tens of thousands or millions sequences per project using Sanger sequencing and therefore the rare microbial groups in studied environment can be easily overlooked (Rastogi and Sani, 2011). Consequently, the limiting factors such as cost and laboriousness and the significant PCR and cloning bias related to the method led to development of new and more powerful technologies for sequencing. Over the last few years the next generation sequencing (NGS) technologies have revolutionized the field of genome and community sequencing, and consequently microbial ecology (Metzker,

2010). NGS technologies typically pursue high throughput by parallelizing the sequencing process, producing thousands to millions of sequences concurrently.

The major advance in NGS is indeed the huge amount of data produced at considerably low cost compared to the Sanger method (Rastogi and Sani, 2011; Metzker, 2010). Projects that took years or months to finish with Sanger sequencing can now be carried out in days, at a fraction of the cost. Divergent features of different NGS technologies, such as read length, throughput and differing error types (homopolymer, insertion/deletion, substitution, random) facilitates the coexisting of multiple platforms in the market (Stahl and Lundeberg, 2012). Decisions as to which NGS platform to use depends on various factors. The drawback of most of these new technologies is their short read length (Stahl and Lundeberg, 2012; Mardis, 2008).

Currently DNA sequencing is one of the most rapidly developing technology of biology. Future advances promise single-molecule sensitivity, uninterrupted real-time sequencing and low cost, but these innovations depend on the development of advanced micro- and nanostructures (Stahl and Lundeberg, 2012). Three major high-throughput sequencing platforms are in use today: the Genome Sequencers from Roche/454 Life Sciences [GS-20 or GS-FLX; (Margulies *et al.*, 2005), the 1G Analyzer from Illumina/Solexa (Bennett *et al.*, 2005) and the SOLiD System from Applied Biosystems. Comparison across the three platforms reveals a trade-off between average sequence read length and the number of DNA molecules that are sequenced. At present, the Solexa and SOLiD systems provide many more sequence reads, but render much shorter read lengths than the 454/Roche Genome Sequencers. This makes the

454/Roche platform appealing for use with barcoding technology, as the enhanced read length facilitates the unambiguous identification of both complex barcodes and sequences of interest.

The 454/Roche Genome Sequencers are called pyrosequencers because their sequencing technology is based on the detection of pyrophosphates released during DNA synthesis (Ahmadian *et al.*, 2006). A few sequencing runs using 454/Roche's pyrosequencing platform can generate sufficient coverage for assembling entire microbial genomes (Hofreuter *et al.*, 2006; Oh *et al.*, 2006), for the discovery, identification and quantitation of small RNAs (Kasschau *et al.*, 2007; Ruby *et al.*, 2006), and for the detection of rare variations in cancers (Thomas *et al.*, 2006), among many other applications. Due to the advantages of the relative long sequence and high throughput, 454 pyrosequencing has been the most popular techniques in the past several years.

Amplicon sequencing has been applied to almost all imaginable environments from deep sea and deep biosphere (Nyyssonen *et. al.*, *in press*) to upper troposphere (DeLeon-Rodriguez *et. al.*, 2013). There are publications about microbial diversity in domestic showerheads (Vornhagen *et. al.*, 2013), human body parts such as belly buttons (Hulcr *et. al.*, 2012), skin (Hanski *et. al.*, 2012) and mouth (Huang *et. al.*, 2011), cow rumen (Fouts *et. al.*, 2012; Mao *et. al.*, 2012), intestines of various animals (Su *et. al.*, 2013; Le Roy *et. al.*, 2012), food products (Nam *et. al.*, 2012; Nieminen *et al.*, 2012) as well as soil, water and sediment (Comeau *et al.*, 2012; Deng *et. al.*, 2012; Siam *et. al.*, 2012; Zhang *et. al.*, 2012).

2.11 Measuring Microbial Diversity

Microbial diversity can be defined by the number of species or different groups (e.g. OTUs) of microbes living in a certain environment, as well as the evenness of the species abundance distribution (Magurran, 2004). In natural environments, microbial communities are typically complex and the diversity is difficult to assess and compare. In order to quantify the diversity as objectively as possible, a variety of diversity indices and richness estimates have been developed and applied (Magurran, 2004). These estimators present the diversity data as a single number that takes various aspects, depending of the indices used, of diversity into consideration. The diversity within individual samples or locations can be assessed using alpha diversity measurements, whereas comparing the community membership and structure between samples or habitats is accomplished by applying beta diversity calculators (Magurran, 2004; Whittaker, 1972 and 1960).

2.11.1 Alpha Diversity

One aspect of alpha diversity is community richness, which describes the number of species present in a certain community (Magurran, 2004). The simplest way of representing the community richness is by the determination of the number of species present a given sampling effort. Different richness estimate calculators, such as Chao1 (Chao, 1984), Abundance Coverage Estimator (Chao and Lee, 1992), jackknife (Burnham and Overton, 1979; Heltshe and Forrester, 1983) and bootstrap (Smith and van Belle, 1984) estimators are applied in measuring the community richness. These nonparametric estimators use the species abundance and occurrence information and

2.11 Measuring Microbial Diversity

Microbial diversity can be defined by the number of species or different groups (e.g. OTUs) of microbes living in a certain environment, as well as the evenness of the species abundance distribution (Magurran, 2004). In natural environments, microbial communities are typically complex and the diversity is difficult to assess and compare. In order to quantify the diversity as objectively as possible, a variety of diversity indices and richness estimates have been developed and applied (Magurran, 2004). These estimators present the diversity data as a single number that takes various aspects, depending of the indices used, of diversity into consideration. The diversity within individual samples or locations can be assessed using alpha diversity measurements, whereas comparing the community membership and structure between samples or habitats is accomplished by applying beta diversity calculators (Magurran, 2004; Whittaker, 1972 and 1960).

2.11.1 Alpha Diversity

One aspect of alpha diversity is community richness, which describes the number of species present in a certain community (Magurran, 2004). The simplest way of representing the community richness is by the determination of the number of species present a given sampling effort. Different richness estimate calculators, such as Chao1 (Chao, 1984), Abundance Coverage Estimator (Chao and Lee, 1992), jackknife (Burnham and Overton, 1979; Heltshe and Forrester, 1983) and bootstrap (Smith and van Belle, 1984) estimators are applied in measuring the community richness. These nonparametric estimators use the species abundance and occurrence information and

nonparametric model to estimate the total number of species present (Hortal *et al.*, 2006). Regardless of sampled environment and organisms of interest, the observed richness is often far lower than the true richness, indicating that more extensive sampling would yield higher number of observed taxa.

In addition to species richness, another important aspect of alpha diversity is community evenness, which is a measure of the evenness in the distribution of species in a sample or environment (Magurran, 2004). The evenness in environmental samples can be represented by Pielou's evenness index or Heip's index of evenness. Pielou's evenness index compares the maximum diversity and estimated diversity derived from Shannon index while Heip's index of evenness measure evenness independently of species richness, working more reliably with communities carrying very low evenness (Magurran, 2004). The concept of community diversity retains both species richness and evenness. A more evenly distributed community is more diverse than a community with few dominant species but the same species richness. Community diversity can be assessed employing a variety of diversity indices including Shannon index and nonparametric Shannon index, Simpson index and inverse Simpson index (1/D) (Magurran, 2004).

These richness estimates and diversity indices, as well as the whole ecological theory, were originally developed for studying macrobiota and not microbes using sequence data and OTUs. It is therefore challenging to assess the suitability of a certain estimate or index for a given dataset and research question, and draw the right conclusions. The relevance of singleton and doubleton OTUs in sequence dataset can easily be

overestimated since they may be products of sequencing or PCR errors (Quince *et al.*, 2011; Kunin *et al.*, 2010; Huse *et al.*, 2010). A more reliable approach for comparing alpha diversity between samples is rarefaction analysis combined with community diversity metrics. The rarefaction analysis compares the samples at a certain sequencing depth, which eliminates the impact of sampling effort. It has been estimated that the Shannon index applied for rarefied data is a fairly reliable measure (Gihring *et al.*, 2012).

2.11.2 Beta Diversity

A major goal in ecological research is recognizing the processes causing spatial variation between communities. This variation is called beta diversity (Whittaker, 1960), which compares the membership and structure of multiple communities and quantifies differences in both taxon composition and relative abundance. As with alpha diversity, beta diversity can be presented using a range of indices. The main measures of β-diversity include Whittaker's measure βW (Whittaker, 1960, 1972), Cody's measure βC (Cody and Diamond, 1975), Routledge's measures βR, βI and βE (Routledge, 1977, 1984) and Wilson and Shmida's measure βT (Wilson and Shmida, 1984). With modern sequencing technologies and large sequence datasets, programs such as LIBHUFF (Singleton *et al.*, 2001), TreeClimber (Schloss and Handelsman, 2006) and UniFrac (Lozupone and Knight, 2005) have been applied in describing beta diversity. Various practical applications are also implemented in popular data analysis pipelines: in Mothur (Schloss *et al.*, 2009) it is possible to compare the membership and structure of multiple communities by creating heatmaps, venndigrams, calculating the

share of overlapping community members as well as drawing dendrograms and calculating the statistical significance of the clustering. Qiime (Caporaso *et al.,* 2010) package contains many similar components. These tools are commonly used in visualizing beta diversity (Magurran, 2004).

2.12 Baricho Water Works

This project was implemented in early 1980s as the first Mombasa and Coastal Water Supply Project funded by World Bank. The main components of the project were as follows: - An intake to abstract raw water from the Sabaki River and a raw water pumping station, treatment works and treated water pumping station to pump water to Mombasa and Kilifi and the construction of Sabaki Pipeline to Mombasa (including a booster pumping station at Jaribuni) and with branches to serve both north and south Kilifi and Mtwapa areas. The World Bank also funded the second Mombasa and Coastal Water Supply Engineering and Rehabilitation Project, which included the development and equipping of wells. There are 8 boreholes which have a daily production potential of 96,000 m^3 of water. However, production is limited to 72,000 m^3 day due to the limitation capacity of the transmission mains. The Sabaki-Nguu Tatu Pipeline has a maximum capacity of 60,000 m^3 day^{-1} while the water demand for Malindi is about 12,000 m^3 day^{-1}. Five pumping sets were installed under the treated water pumping and have been numbered one, two, three, four and five while the old treated water pump is numbered as six. However all these highlift pump sets are not operational except for No's one, two and six.

Water from Baricho Water Works is pumped through the Sabaki pipeline to Nguu-Tatu Reservoirs of capacity 18,000 m^3 which are located some 10 Km from Mombasa. The Sabaki Pipeline is 110 Km long and is made of bitumen varnished ductile iron pipes manufactured in West Germany. The size of the pipeline ranges from 600 to 800 mm in diameter. A balancing tank is located at Lower Ribe terminating the pumping section and with a bypass pipeline. From the reservoirs, water is conveyed to Mombasa North Mainland through a 20" diameter gravity trunk line to Kisauni. A second trunk main of 28" diameter conveys water by gravity across the Nyali Bridge to Mombasa Island. Under the Second Mombasa and Coastal water supply Project, a rehabilitation works and a bypass to Lower Ribe tanks was commissioned to pump water directly to Nguu-Tatu reservoirs. At the same time, a section of 10 km long from Lower Ribe was duplicated to reduce head losses.

2.13 Mzima Pipeline System

Mzima Springs are located south west of the Chyulu Hills in Tsavo West National Park (Appendix I). The springs have been gauged since 1951 and shown a flow variation between 2.6 $m^3 s^{-1}$ (225,000 $m^3 day^{-1}$) and 5.9 $m^3 s^{-1}$ (510,000 $m^3 day^{-1}$) with a mean of 3.5 $m^3 s^{-1}$ (302,000 $m^3 day^{-1}$). The current abstraction is about 0.4 $m^3 s^{-1}$ (35,000 $m^3 day^{-1}$). Water from the springs flows through three large pools: Hippo pool, Long pool and Chalk beach pool before discharging into the Mzima River, which is 4 km further downstream. The existing source works comprises a 670 m long sheet pile cut-off wall with infiltration trench/gallery. Depending on groundwater table level, between 0.8 and 1.1 $m^3 s^{-1}$ is abstracted by this cut-off and conveyed to an overflow chamber at the end

of an initial conveyance section comprising some 1200 metres of 30" diameter pipe work. At this point all but about 0.4 m^3 s^{-1} overflows into the long pool about half way along its length. From this overflow chamber, the conveyance section continues for a further 1,880 metres (110 metres in a tunnel of 42 inches wide x 72 inches deep and the balance in 48 inches diameter pipe work) before the Head works of the Mzima Pipeline proper. At the head works, which is some 3 km away from the source works, there is a provision for a duplication of the existing pipeline (Mzima II). Water from the springs is believed to possess excellent chemical, physical and biological characteristics, which do not require any particular form of treatment except for disinfection. However, as a precaution, water is usually disinfected. Tropical Chloride of Lime or High Test Hypochlorite (HTH) is used as the disinfectant.

The construction of the pre-stressed concrete Mzima pipeline started in 1953 and reached Mazeras in 1957 and is 220 km long. The diameter of the pipeline ranges from 36 inches to 21 inches and has been in operation for 50 years. Certain sections of the pipeline are prone to bursts or leakages and due to this about 15 km of the pipeline have already been replaced by steel pipes. The springs are located at an altitude of 2230 feet (680 metres) above sea level and therefore water flows by gravity to Mombasa. The pressures along the pipeline are controlled by 10 Break Pressure Tanks (BPT), serially numbered 1 to 10 in the direction of flow of water, before discharging into the reservoirs at Mazeras (81,000 m^3). The principle branch lines are at Voi and Mariakani. Altogether, there are an estimated 300 off-takes (varying in diameter from 12 mm to 250 mm) which supply water to a 10 km wide corridor of the pipeline. From the

Mazeras reservoirs, water gravitates to six (6 No.) reservoirs situated at Changamwe with total capacity 6,600,000 gallons before water flows by gravity to Mombasa Island. Immediately after Mazeras tanks, there is a take-off which supplies water by pumping to Kaloleni Trading Centre, including all the consumers en-route the pipeline strip. The pipeline is about 22 km long and is 200 mm in diameter, (Asbestos and concrete) which was commissioned in 1969.

CHAPTER 3: MATERIALS AND METHODS

3.1 Study Area

The study focused on Mombasa County which is part of the Coastal Kenya region bodering the Indian Ocean. The region is supplied with water from five bulk supply systems, which currently provide an estimated total of 115,000 m^3 day^{-1} of water. The bulk water sources include the Baricho Water Works, Mzima Pipeline System, Marere Pipeline System, Tiwi Boreholes and Taveta-Lumi Water System, which provide water at rates of 60,000 m^3 day^{-1}, 35,000 m^3 day^{-1}, 7,000 m^3 day^{-1}, 10,000 m^3 day^{-1}, and 3,000 m^3 day^{-1} respectively. The present total water demand is 250,000m^3 day^{-1}.

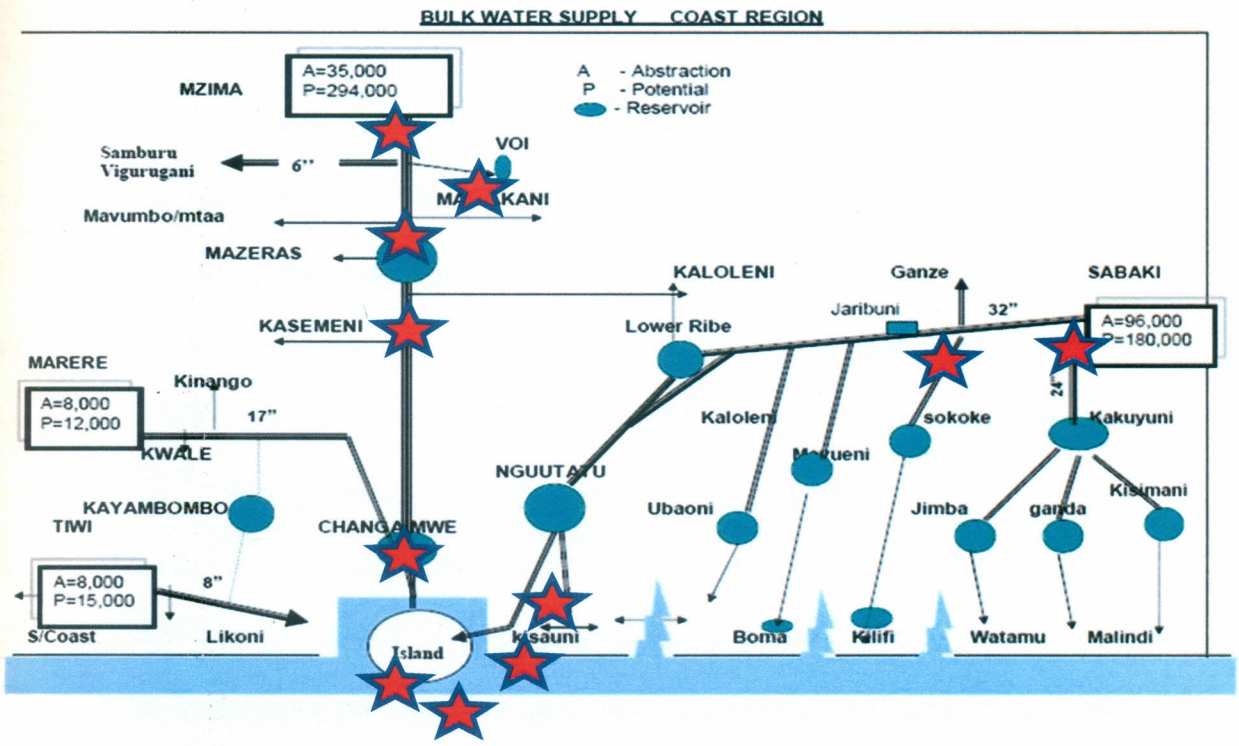


Fig. 3.1 Schema of water distribution systems supplying water to Mombasa County showing the points sampled

Key: The red stars indicate the sampling points while the arrows indicate distribution lines for communities supply.

3.2 The Sampling Points

Two of the three water distribution networks supplying water to the Mombasa County, the Mzima and Baricho waterworks, were studied. Samples were collected from the water sources, that is Mzima Springs intake and the Sabaki River boreholed aquifer, and along the distribution lines to user point meters, both in the sparsely populated areas of Mazeras and the densely populated city estates of Kibokoni and Mnazi moja (Fig. 3.1) The sampling points included the initial water intakes, chlorination tanks, break pressure tanks, main pipes, T- junctions and household meters. These points were purposively selected for sampling to deliberately include key representations of the distribution system.

3.3 Sampling

Water sampling was carried out thrice; May 2013, September 2013 and March 2014, so as to obtain samples during the dry, long and short rain seasons. Twenty one samples (13 water and 8 biofilms), were collected at each sampling points using clean two liter sterile glass stoppered bottles giving a total of sixty three samples over the three sampling episodes. Biofilm samples were collected by scrapping from the internal pipe surfaces accessed through break pressure and chlorination tank openings, using a sterile scoop and bottled in 100 ml of sterile water. The areas of pipe from which the samples were scrapped were from 3 to 6 cm^2, based on site accessibility and shear pressure. The samples were kept in airtight bottles in an icebox upon collection, for delivery to the laboratory.

3.4 Determination of Physico-chemical Quality of Water

In establishing the physical and chemical quality of water in the distribution systems: pH, temperature, residual chorine, nutrients and selected metals were measured. Nutrients levels were estimated as nitrate nitrogen and phosphate phosphorus using the Palintest ® meter which applies the photometer method. To determine the concentration of phosphate phosphorus in water samples, 2.0 ml of water sample was pipetted in to a Phosphate/12P Tube. One Tubetests Phos No 1 tablet was added, crushed and mixed to dissolve. This was followed by the addition of one Tubetests Phos No 2 tablet, which was crushed and mixed to dissolve. The tube containing the resultant mixture was gently inverted several times to mix and left to stand for 10 minutes to allow colour development. Color intensity was was determined by reading light extinction using a Palintest Photometer (phot 91) which gives extinction values in mg L^{-1} PO_4.-P

Nitrate nitrogen (NO_3.-N) levels were determined using the nitrate nitrogen test kit of the Palin test meter. In the Palintest Nitratest method, nitrates are first reduced to nitrites and the resulting nitrites determined by a diazonium reaction method. The reduction stage was carried out using the unique zinc-based Nitratest Powder, and Nitratest Tablet, which aids rapid flocculation after the one minute contact period. The nitrites resulting from the reduction stage are coupled with sulphanilic acid in the presence of N-(1-naphthyl)-ethylene diamine to form a reddish dye. The reagents are provided in a single Nitricol tablet which is simply added to the test solution. The intensity of the colour produced in the test is proportional to the nitrate concentration

To each Nitratest Tubes, water samples were added up to the 20 ml mark. A level spoonful of Nitratest Powder and one Nitratest tablet were added, the tubes shaken and allowed to stand for one minute. The tubes were next gently inverted three or four times to aid flocculation and left to stand for three minutes to ensure complete settlement. The clear solution was decanted into a round test tube to the 10 ml mark. A Nitricol tablet was added crushed and mixed to dissolve then allowed to stand for 10 minutes to allow full colour development. Color imtensity was was determined by reading light extinction using a Palintest Photometer phot 63. The photometer gives extinction values in mg L^{-1} NO_3-N.

Levels of iron, manganese, lead and zinc were determined using the atomic absorption mass spectrophotometry (Aurora AI 1200, Vancouver Canada). Analytical grade reagents and double distilled water were used throughout the study. All glassware and plastic containers used were washed with detergent solution followed by 20% nitric acid and then rinsed with tap water and finally distilled water. Lead nitrate, iron chloride, manganese chloride and zinc chloride were used for preparation of lead, iron, manganese and zinc standards respectively. To ensure the removal of organic impurities from the samples and therefore prevent the interference in analysis, the samples were digested with concentrated nitric acid. An amount of 5 mL of concentrated HNO_3 was added to 100 mL of a water sample in a 250 mL conical flask then heated on a hot plate and evaporated until 20 mL was left. After cooling, another 5 mL of concentrated HNO_3 was added and the digestion continued until 10 mL of sample was left. The digestate

was next filtered and diluted with distilled water to a volume of 100 mL and stored in the refrigerator.

The stock solution of lead, iron, manganese and zinc were prepared by dissolving 24.62, 4.84, 3.91 and 1.60 g of lead nitrate, iron chloride, manganese chloride and zinc chloride respectively in 68% of nitric acid in a one litre volumetric flask. Each mixture was shaken and the flask filled up to the 1 L mark with the nitric acid solution for each metal. Calibration solutions of the target metal ions were prepared from the standard stock by serial dilution. The digested water samples were analyzed for the presence of lead, iron, manganese and zinc using Aurora AI 1200, Vancouver Canada Atomic Absorption Spectrophotometer. Standards of known metal ion concentration were subjected to same treatment as the water samples and used to determine actual metal ion concentration. The wavelengths for the determination of lead, iron, manganese and zinc were 283.31, 248.33, 279.48 and 213.9 nm respectively. The digested samples were analyzed in triplicates with the average concentration of metals being read in mg L^{-1} by the instruments after extrapolation from the standard curve.

3.5 Characterization of the Bacterial Community

3.5.1 DNA Extraction

Two liters of water were filtered through 0.45 μm filter polycarbonate membranes (Milipore, MA) to concentrate suspended bacteria for subsequent DNA extraction. The bacteria on the polycarbonate membrane were re-suspended in water in a 1.5 ml

microcentrifuge tube, vibrated for 10 minutes and finally collected by centrifugation at 10000 xg for 5 min for DNA extraction. Extraction followed the phenol/chloroform method (Sambrook *et al.,* 1989). The cultures were transferred into 1.5 mL of eppendorf tubes, centrifuged at 13000 xg for five minutes and the supernatants discarded. The cell pellets were re-suspended in 200 µL of solution A (50 mMTris pH 8.5, 50 mM EDTA pH 8.0 and 25 % sucrose solution). To this 5µL of Lysozyme (20 mg mL^{-1}) and 5 µL of RNAse A (20 mg mL^{-1}) were added, gently mixed and incubated at 37 °C for 1 hour. Following incubation, 600 µL of solution B (10 mMTris pH 8.5, 5 mM EDTA pH 8.0 and 1 % SDS) was added and contents mixed by inverting the eppendorfs several times. 10 µL of Proteinase K (20 mg mL^{-1}) was then added, mixed gently and incubated at 50 °C for 1 hour. Presence of DNA was checked on 1 % agarose and visualization under ultraviolet of ethidium bromide stained extracts. The extracted genomic DNA was used as a template for subsequent PCR amplification.

3.5.2 PCR Amplification

The twenty one DNA samples extracted were independently amplified using paired primer sets, 338F (5'-ACTCCTACGGGAGGCAGCAG-3') and 533R (5'-TTACCGCGGCTGCTGGCAC-3') (Huse *et al.*, 2008) targeting the V3 hypervariable region of the 16S RNA gene. Polymerase chain reaction (PCR) mixes contained 17.1 µL of sterile water, 2.5 µL 10X of reaction buffer (Sigma), 2.5 µL of each deoxyribonucleotide triphosphate (dNTP 2.0 µM), 0.5 µL of each primer (10 µM), 0.4 µL of DNA polymerase (High Fidelity Taq, Roche) and 2 µL of DNA template in a final volume of 25 µL. The DNA was amplified using a T3000 thermal cycler

(Biometra, Göttingen, DE). The following conditions were used; initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 45 s, extension at 72 °C for 1 min and a final extension at 72 °C for 7 min (Zhang *et al.*, 2009). The PCR products obtained were purified with the Agencourt® AMPure® Kit (Agencourt Bioscience Corporation, MA, USA). The quality of these samples was assessed using 1.5% agarose gel, and evaluation of the AD260/280 ratio calculated using the ND-1000 Spectrophotometer NanoDrop® (Thermo Scientific, Wilmington, DE) The different amplicons obtained from each studied plot were pooled before sequencing.

3.6 454 Sequencing

Pyrosequencing of the 16S rRNA gene amplicon libraries was performed to characterize the bacterial diversity in drinking water following whole community DNA extraction and purification. For each DNA sample, amplicon libraries were generated with primers targeting the V3 hypervariable region of the 16S rRNA gene: 338F (5'-ACTCCTACGGGAGGCAGCAG-3') and 533R (5'- TTACCGCGGCTGCTGGCAC-3') (Huse *et al.*, 2008). To pool multiple samples for one run of 454 sequencing, barcode sequences unique to each sample were attached to the primers (Hamady *et al.*, 2008). Polymerase Chain Reaction (PCR) amplification was performed with the FastStart High Fidelity PCR system in a total volume of 50 µL, containing 5 µL of FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$, 4% DMSO, 200 µM dNTPs, 0.4 µM forward and reverse primers, 10-100 ng of DNA, and 2.5 U FastStart High Fidelity Enzyme Blend (Roche Diagnostics, Germany). All amplicon products from different samples were mixed in equal concentrations and purified using

Agencourt Ampure beads (Agencourt Bioscience Corporation, MA, USA). The purified samples were sequenced utilizing Roche 454 FLX titanium instruments and reagents following manufacturer's guidelines.

3.7 Bacterial Community Analysis

The 454 datasets were processed with the QIIME (Quantitative Insights Into Microbial Ecology) (Caporaso *et al.,* 2010), which is an open-source metagenomics pipeline designed to provide self-contained microbial community analyses. It integrates commonly used third-party tools and implements many diversity metrics, statistical methods, and visualization tools for analyzing microbial data.

3.7.1 Upstream Analysis

Sequences were depleted of barcodes and primers (Appendix II). Short sequences of < 200bp, sequences with ambiguous base calls, and sequences with homopolymer runs exceeding 6bp were also removed. The remaining sequences were then denoised and operational taxonomic units defined by clustering at 3% divergence (97% similarity) followed by removal of singleton sequences and chimeras using the Uclust algorithm.

The cleaned datasets were clustered into molecular Operational Taxonomic Units (OTUs) with a 97% identity threshold, using the Uchime algorithm (Edgar *et al.,* 2011), which is integrated in the Usearch 6.1 (Edgar, 2010) pipeline and implements a native *de novo* chimera identification and removal feature. The longest sequence from each cluster was selected as the OTU representative sequence to be adopted for use in

taxonomic identification of the OTU. Assigning taxonomic ranks and nomenclature to each sequence of the representative set was carried out by querying with RDP classifier (Wang *et al.,* 2007) with a confidence level of 0.8 against the Silva database (Quast *et al.,* 2013) which includes microbial eukaryotes in order to infer the likely functional roles for members of the community. PyNAST (Caporaso, Bittinger, *et. al.*, 2010) was used to align the sequences against a template sequence alignment.

A phylogenetic tree representing the relationships among sequences was inferred from the multiple sequence alignment generated in the heatmap function of Biodiversity R using Phyloseq and the Neighbor Joining algorithms in GGplot. The stage upstream in QIIME involved constructing an OTU table using the Genomics Standards Consortium candidate standard Biological Observation Matrix (BIOM) format (McDonald, Clemente, *et. al.*, 2012).

3.7.2 Downstream Analysis

A second level of quality-filtering based on OTU abundance was applied, discarding OTUs with a number of sequences <0.005% of the total number of sequences (Bokulich *et al.,* 2013). This step greatly reduced the problem of spurious OTUs, most of which were present at very low abundance. Taxa summaries indicating relative abundance in samples on multiple taxonomic levels were generated. These were used to identify outliers, visually identify expected patterns and differences between and among samples.

3.7.2.1 Alpha-diversity Analysis

Alpha-diversity is the diversity of organisms in one sample or environment. QIIME was used to generate plots showing alpha-diversity at different rarefaction levels. Rarefaction curves were used for assessing the sequencing effort, representing and comparing the microbial communities. The resulting images were used to estimate the true species richness and evenness of communities. Phylogenetic distance (PD) (Faith, 1992) was used to generate an interactive HTML document with figures showing the results for each alpha-diversity metric and for each group of samples. The diversity was estimated using Shannon, Simpson and Inverse Simpson indices for diversity comparisons determined as the ratio of the number of OTUs shared and the total number of OTUs in the two water line samples.

3.7.2.2 Beta-diversity Analysis

Beta-diversity is the difference in diversities across samples or environments. The bacterial assemblages retrieved in the different samples were compared by means of Principal Components Analysis (PCA) (Chen *et al.*, 2009a). The beta-diversity distance matrices were analyzed using UPGMA. OTU heatmaps were generated using the phyloseq package for R (McMurdie and Holmes, 2013) for exploratory analysis of microbiomes by mapping abundance values to a color scale in a condensed, pattern-rich format, in which each row corresponds to an OTU and each column corresponds to a sample. Nonmetric multidimensional scaling (NMDS) of the Bray–Curtis distance was used to determine the order of the OTUs and samples.

3.7.3 Statistical Analysis

Several visual and statistical tests were conducted. Distance histograms were used to compare distances graphically. The diversity indices were also used to calculate Mantel statistics (correlation coefficients) between bacterial community and sampling site matrices. The correlation between OTU abundance and levels of selected physico-chemical properties was determined using Pearson's correlation test. Multivariate analyses were performed to explore relationships between the beta-diversity distance matrix or covariates.

Analysis of variance (ANOVA) was used to compare the relative abundance of each taxon in the OTU table of the two water lines and sample types. Analysis of similarities (ANOSIM) and redundancy analyses were employed to test variation within and between water distribution systems and sample types. The Kruskal-Wallis test was used to determine whether the bacterial composition and diversity significantly differed between distribution systems.

CHAPTER 4: RESULTS

4.1 Bacterial Community Composition and Diversity in the Water Distribution Systems

The 454-FLX Titanium run generated a total of 27,937 sequences after quality filtering and multiplex binning from the 21 samples. After quality checking of V4 region of the 16s RNA, reads were denovo clustered into operational taxonomic units (OTUs) based on their sequence similarity giving rise to 2,294 unique OTUs.

The rarefaction analysis of 454-FLX amplicon data sequences indicates the sampling effort adequately captures the entire diversity of the bacterial community (Fig. 4.1). The rarefaction curve for all samples approaches an asymptote indicating proper sampling of hyper diverse communities. The composite rarefaction curve of all treatments follows roughly the same trajectory as individual treatment curves. This suggests that if one were to sample more deeply within a given treatment, new OTUs would be encountered less often.

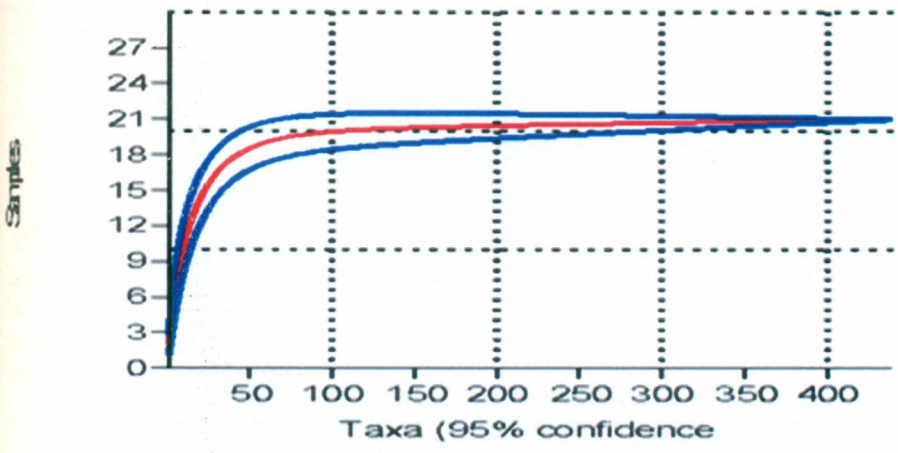


Fig. 4.1 Sample rarefaction curves for the microbial community in water distribution systems.

A total of 2,294 OTUs were detected in the water distribution systems of Mzima and Baricho. From Mzima water distribution systems, 424 OTUs were detected in the water with 372 being unique to Mzima water only while 419 OTUs were identified from biofilms collected from the Mzima pipeline with 348 being unique to these biofilms. A total of 40 OTUs were found to be common in both Mzima water and its pipeline biofilms. Baricho water distribution system yielded 635 and 587 OTUs in its water and biofilms respectively, out of which 587 and 534 were unique to the water and biofilms respectively. A total of 22 OTUs were detected in both Baricho water and biofilm samples. Only three OTUs were common to the water and biofilms of both Mzima and Baricho water distribution systems (Fig. 4.2).

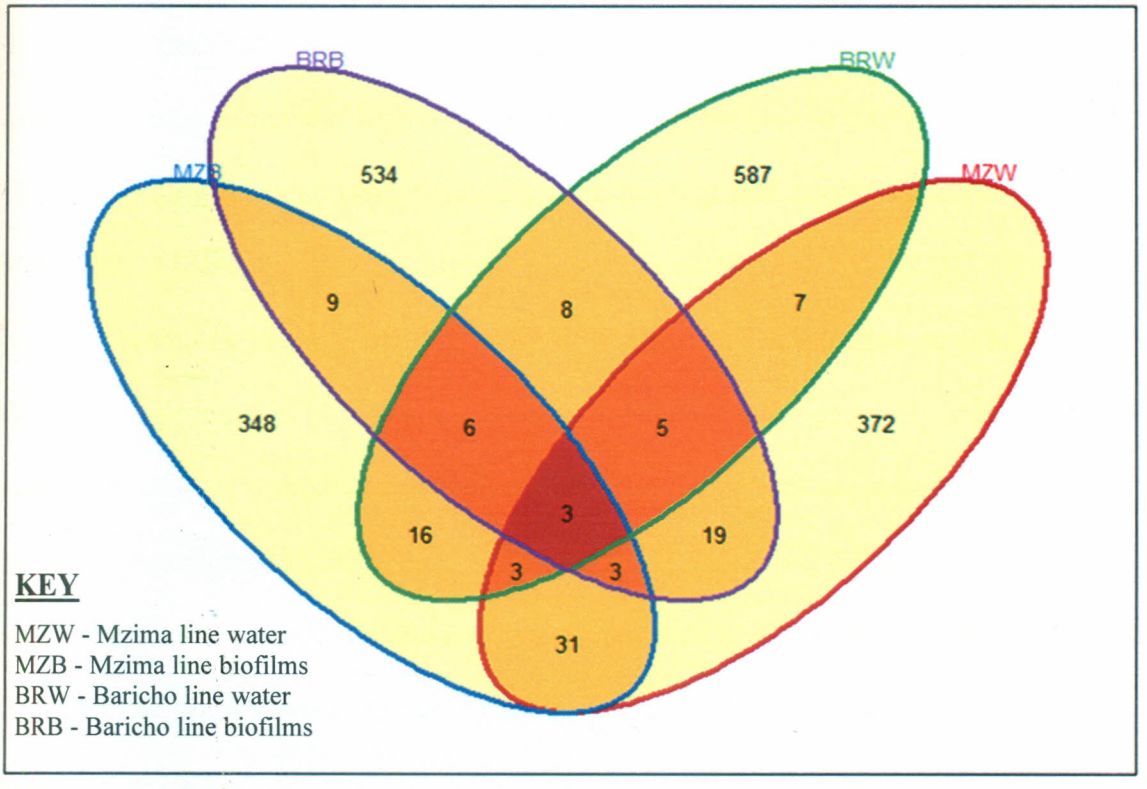
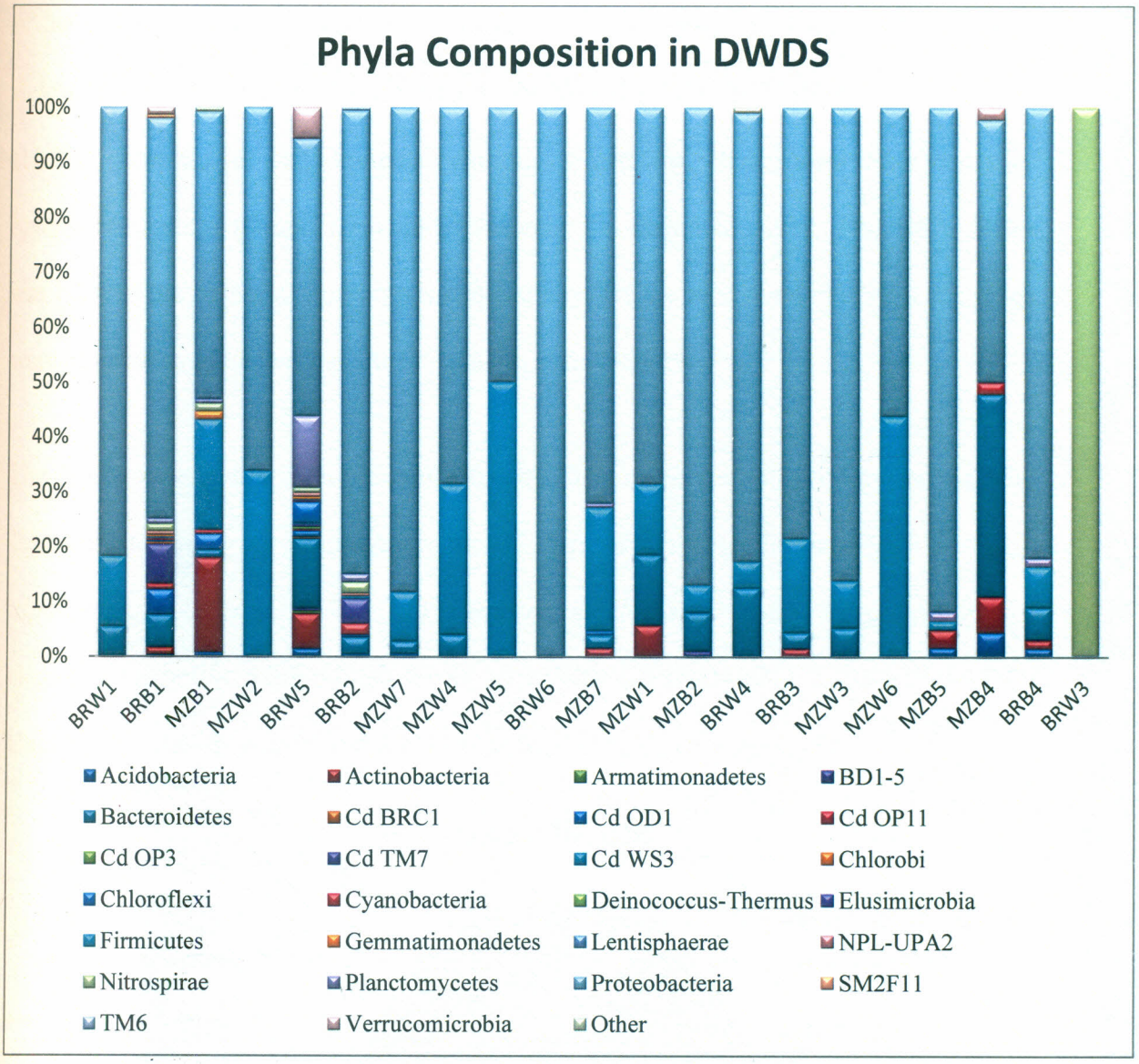


Fig. 4.2 Venn diagram representing shared and unique bacteria OTUs at 97% sequence similarity between Water distribution systems and sample types

A total of 20 bacterial phyla and 6 candidate phyla were identified from pooled samples. The bacteria phyla were dominated by the *Proteobacteria* (73.2% of total bacterial genomes), followed in abundance by *Firmicutes* (13.4%), *Bacteriodetes* (5.9%), *Actinobacteria* at 2.3% and *Planctomycetes* (1.2%). Candidate phyla accounted for 0.9% of the total bacterial genomes (Fig. 4.3). Candidate divisions are a lineage of bacteria, the existence of which until recently was known solely through environmental 16S rRNA sequences and for which no cultured representatives have been found as is the requirement for taxonomy. There was no successful blast hit for 0.1% of the OTUs.

The phylum *Proteobacteria* consistently contributed the highest proportion of phylotypes in all water samples making up all (100%) bacteria phylotypes in BRW6. The lowest representation of 47.8% of the phylotypes was recorded in MZB4. *Firmicutes* contributed the highest proportion of phylotypes in MZW5 at 50%, with BRB1 recording 0.3% of phylotypes from the phylum and none recovered in BRW6 and MZB4. MZB4 recorded the highest proportion of *Bacteroidetes* at 37% with none being recovered in MZW2, MZW5, MZW6, MZB5 and BRW6 (See Appendix III for the absolute proportion phylotypes). The *Actinobacteria* were generally more common in biofilm samples compared to water samples and had a 17% representation in MZB1 but was not recovered in MZW2, 3, 4, 5, 6 and 7, BRW1, 4 and 6 BRB2 and BZB2. All sequences denoting candidate divisions were identified in BRW5 and BRB1, while *CD TM7* and *CD OD1* accounting for 7.3% and 4.6% of the phylotypes respectively were recovered in BRB1.



KEY

MZW- Mzima line water
MZB- Mzima line biofilms
BRW- Baricho line water
BRB- Baricho line biofilms

Fig. 4.3 Taxonomic affiliation of metagenomic reads obtained from the Mzima and Baricho water distribution systems of Mombasa County

Results for complete datasets were evaluated by BLASTX analysis against SILVA 119 reference database using QIIME v 1.8.0 software

Considering the class level distributions of phylotypes in the two water lines, a total 68 classes were recovered at two water distribution systems. The dominant class in terms

of the percentage contribution of phylotypes was *Gammaproteobacteria* (38%) with *Betaproteobacteria* (21.6%), Fimircute *Bacilli* (11.4%) and *Alphaproteobacteria* (11.2%) occurring at subdominant positions. *Clostridia*, *Bacteroides Flavobacteria, Sphingobacteria, Actinobacteria* and *Cytophagia* contributed 1.9%, 3%, 1.6%, 2.0% and 1.2% of the bacterial phylotypes respectively. Of the recovered classes 0.9% reads were of uncultured bacteria and 0.1% generated no blast hits.

The Baricho water line was predominated by the *Proteobacteria* phylotypes with the *Gammaproteobacteria* dominating at 37.5% of the phylotypes followed by *Alphaproteobacteria* at 16.7% and *Betaproteobacteria* at 13.8% (Fig. 4.4). Other classes occurring at subdominant levels were *Nitrospirae* (11.6%) and *Bacilli* (4.8%). Unidentified classes constituted 2.7% of the total OTUs in the water line.

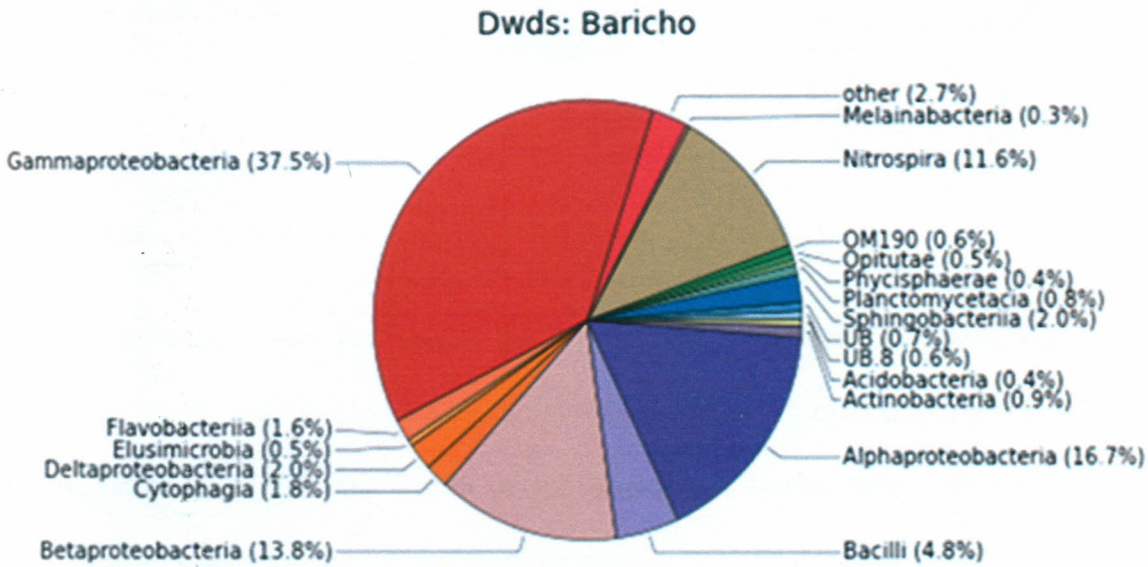


Fig. 4.4 Class level bacteria distribution in Baricho water line

The Mzima water line on the other hand, had *Gammaproteobacteria* phylotypes dominating at 39.1% followed by *Betaproteobacteria* at 24.0% and *Bacilli* at 16.3%. *Alphaproteobacteria* contributed only 5.6% of the phylotypes, with other subdominant classes being the *Flavobacteria* (4.0%) and *Clostridia* at 3.1% (Fig. 4.5). Unidentified classes constituted only 0.7% of the total OTUs in Mzima water line.

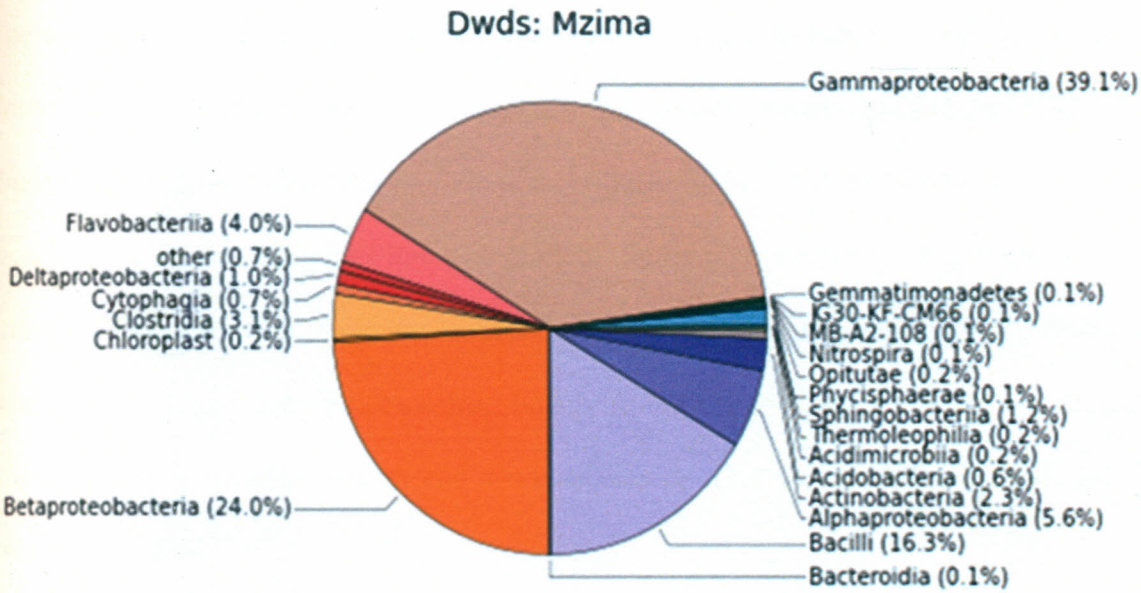


Fig. 4.5 Class level distribution in Mzima water line

At the ordinal level of taxa distribution, the *Pseudomanadales* were the dominant group contributing 24.9% of the phylotypes. Subdominant groups included *Burkhoderiales* and *Bacilliates* at 12.7% and 10.3% of the phylotypes respectively. Other notable genera were *Neisseriales* (5.2%) and *Enteriobacteriales* (3.0%) which are made up of critical human pathogens. *Rhizobiales* which are usually soil inhabitants and plant symbionts contributed 4.6% of the 136 identified order phylotypes.

A total 246 families were identified from all the analysed samples, and found to be dominated by *Moraxellallaceae* (20%) followed by *Comamodaceae* (8.2%). Identified families known to be sources of pathogenic bacteria were *Bacillaceae* (5.6%), *Neisseriaceae* (5.2%), *Pseudomonaceae* (4.9%), *Enterobacteriaceae* (3%), *Aeromonadaceae* 1.7% and *Clostridiaceae* (1%).

The total number of genera recovered was 498, the most abundant being the genus *Acinetobacter* (19.5%) which was recorded in all samples except MZW2 and BRB2. *Rheinheimera, Bacillus* and *Pseudomonas* contributed 5.8%, 5.5% and 4.6% of the total genera respectively. Samples from the Baricho water line evidently displayed more diverse generic composition comparable to Mzima water line samples (Fig. 4.6). Other genera with high occurrence were *Vogesella* (4.6%), *Flavobacterium* (2.8%), *Exignobacterium* (2.4%), uncultured *Comamodaceae* (2.2%), *Massilia* (2.1%) and *Rhizobium* (1.6%).

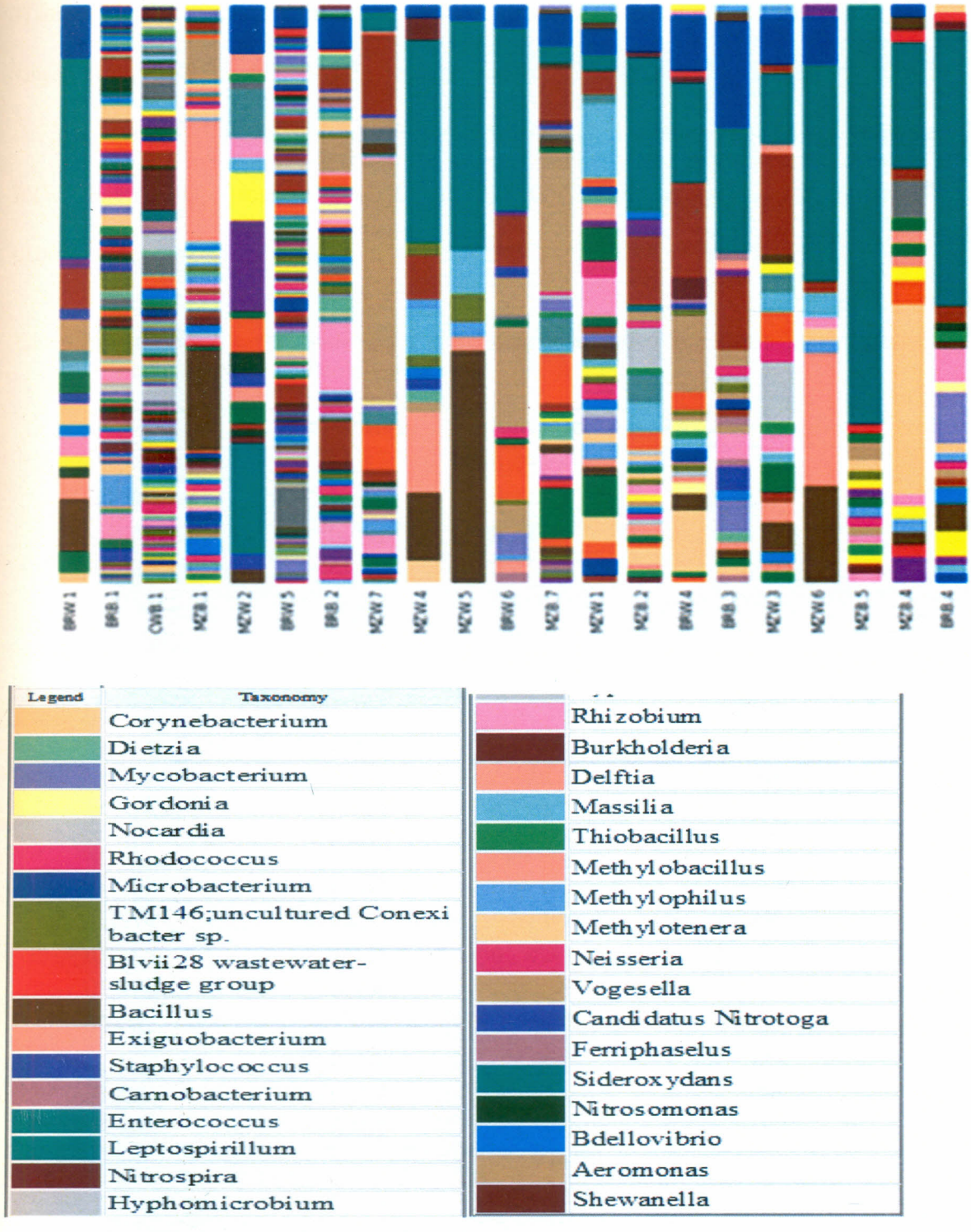


Fig. 4.6 Generic levels bacterial abudance in the drinking water distribution systems

The Shannon index of diversity (H′) was determined for all samples at phylum level (Table 4.1). The Shannon index of each sampling site at a genetic distance of 3% ranged from 0.00 to 1.725. The highest bacterial diversity was found in BRW5 (1.725), followed by MZB1 (1.391) and MZB4 (1.201). Simpson indices indicate BRW5 having the greatest diversity (0.702) and hardly any diversity in BRW6 (0.000).

Table 4.1 Summary diversity indices for each water distribution system samples at phylum level using 97% similarity values

Sites	Shannon	Simpson	Inverse Simpson
MZW1	0.950	0.493	1.97
MZW2	0.639	0.447	1.81
MZW3	0.492	0.247	1.33
MZW4	0.740	0.452	1.82
MZW5	0.693	0.500	2.00
MZW6	0.685	0.492	1.97
MZW7	0.440	0.212	1.27
MZB1	1.391	0.652	2.88
MZB2	0.503	0.236	1.31
MZB4	1.201	0.628	2.68
MZB5	0.388	0.153	1.18
MZB7	0.799	0.429	1.75
BRW1	0.585	0.311	1.45
BRW4	0.615	0.314	1.46
BRW5	1.725	0.702	3.35
BRW6	0.000	0.000	1.00
BRB1	1.173	0.456	1.84
BRB2	0.739	0.282	1.39
BRB3	0.654	0.352	1.54
BRB4	0.712	0.316	1.46

Shannon diversity indices at the genus level (Table 4.2) ranged from 4.54 in BRB1 to 2.65 in MZW5 and MZB5 with the highest diversity being displayed by Baricho biofilms at 3.825±0.618. The Simpson indices (expressed a 1 – D) for all 454-FLX amplicon communities at genus level (Table 4.2) were close to 1, indicating a very high heterogeneity at each site. Simpson diversity index was highest in BRW5 (0.982) and lowest in BRW6 (0.916), both values being close to the maximum possible value of 1.

Table 4.2 A summary of diversity values in each water distribution system sample at genus level using 97% similarity values

Sites	Shannon	Simpson	Inverse Simpson
MZW1	3.84	0.961	25.86
MZW2	3.05	0.928	13.92
MZW3	3.27	0.935	15.39
MZW4	3.16	0.938	16.10
MZW5	2.65	0.919	12.30
MZW6	2.84	0.927	13.73
MZW7	3.11	0.920	12.56
MZB1	4.15	0.966	29.38
MZB2	3.48	0.938	16.25
MZB4	3.57	0.956	22.76
MZB5	2.65	0.879	8.28
MZB7	3.54	0.948	19.14
BRW1	3.33	0.935	15.34
BRW4	3.38	0.938	16.25
BRW5	5.10	0.982	54.61
BRW6	2.88	0.916	11.87
BRB1	4.54	0.965	28.23
BRB2	4.13	0.952	20.81
BRB3	3.42	0.938	16.09
BRB4	3.21	0.926	13.51

4.2 Relative Abundance of Bacterial Communities in Water Distributions Systems

Based on the Chao richness estimator, an overall phyla richness of approximately 32.6 and a total abundance 2331 were computed. The sample with highest phylum richness was BRW5 (20) and lowest number was 1 in BRW6. Other sites with high phylum richness were BRB1 (17), BRB2 and MZB1 both with a richness of 11 (Table 4.3). These sites had a correspondingly high abundance with BRW5 recording the highest abundance value of 423 followed by BRB1 (303) BRB2 (200) MZW7 (179) and MZB1 (139). The lowest number of phyla abundance was 40 in MZW5. The highest recorded evenness was in MZW5 and BRW6, while the lowest was in BRB 1 and BRB2.

Table 4.3 Taxa richness, evenness, and abundance in the different samples of each water distribution systems at phylum level using 97% similarity values

Sites	Richness	Abundance	Evenness
MZW1	4	70	0.646
MZW2	2	83	0.948
MZW3	3	58	0.545
MZW4	3	51	0.699
MZW5	2	40	1.000
MZW6	2	48	0.992
MZW7	4	179	0.388
MZB1	11	139	0.366
MZB2	4	100	0.413
MZB4	6	46	0.554
MZB5	5	62	0.295
MZB7	6	129	0.370
BRW1	3	55	0.598
BRW4	4	104	0.463
BRW5	20	423	0.281
BRW6	1	104	1.000
BRB1	17	303	0.190
BRB2	11	200	0.190
BRB3	4	70	0.481
BRB4	6	67	0.340

Using the chao richness estimator, the total richness of bacterial genera was estimated to be approximately 673 with an abundance of 11655 genera recoveries from samples of both water distribution lines. Samples from Mzima and Baricho water lines had a mean richness of 71.33±42.58 and 155.75±147.121 respectively. The sample with highest phylum richness was BRW5 (453) and lowest number of different genera was 22 in MZW5 (Table 4.4). Other sites with high genera richness were BRB1 (290), BRB2 (189) and MZB1 (185). These sites had a correspondingly high abundance BRW5

highest with 2115 followed by BRB1 (1515) BRB2 (1000) MZW7 (895) and MZB1 (695). The lowest number of genera was 200 in MZW5.

Table 4.4 Estimated richness, evenness, and abundance of genera for each sample of the water distribution systems investigated using a 97% similarity value

Sites	Richness	Abundance	Eevenness
MZW1	86	350	0.543
MZW2	45	415	0.467
MZW3	56	290	0.467
MZW4	45	255	0.523
MZW5	22	200	0.644
MZW6	31	240	0.553
MZW7	78	895	0.286
MZB1	185	695	0.344
MZB2	84	500	0.387
MZB4	69	230	0.514
MZB5	57	310	0.248
MZB7	98	645	0.354
BRW1	61	275	0.459
BRW4	71	520	0.416
BRW5	453	2115	0.361
BRW6	39	520	0.459
BRB1	290	1515	0.323
BRB2	189	1000	0.328
BRB3	76	350	0.403
BRB4	67	335	0.370

Rényi diversity profiles curves, which order samples on the basis of diversity and evenness from the lowest to highest diversity, revealed that all sampling sites have profiles that decline from left to right. This indicates that species are not evenly distributed for all the sampled sites (Fig. 4.7). The starting position at the left-hand side of the profile is an indication of the species richness and these profiles start at a higher

level indicating that they have higher richness. Since some of the profiles intersect, it is not possible to order the sites from lowest to highest diversity. On a general scale all sampled points have higher species richness, but lower species evenness.

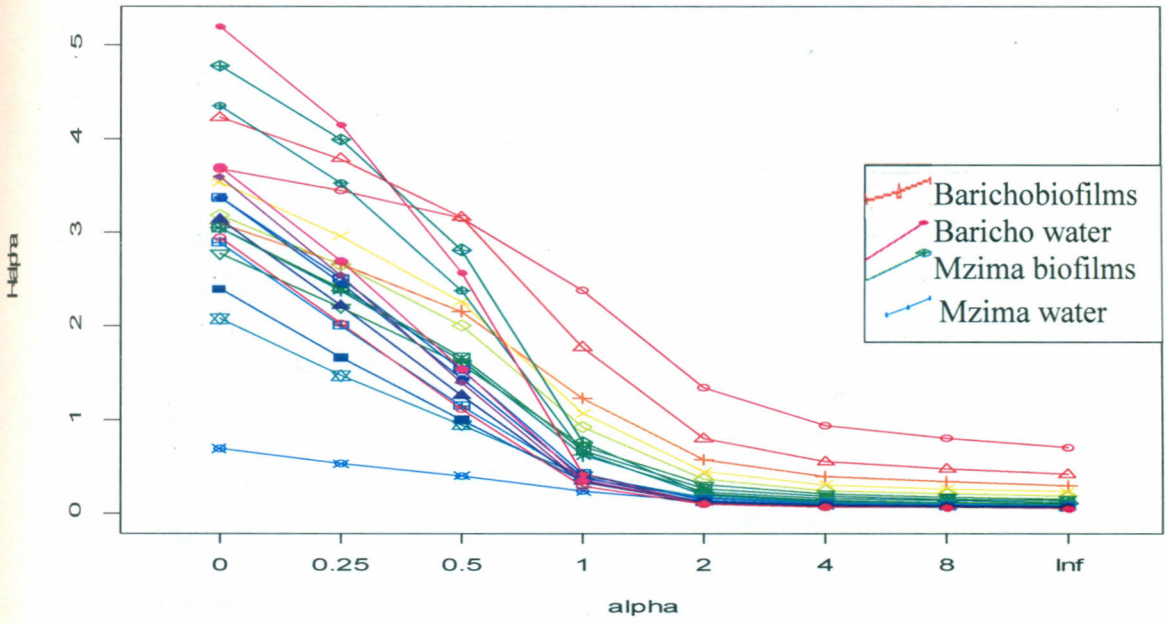


Fig. 4.7 Rényi diversity profiles on overall taxa richness and evenness of water distribution systems.

Non-metric multi dimension scaling ordination of class phylotype matrix maximizing the rank order correlation, yielded a two dimensional solution (final stress = 0.2009464) (Fig. 4.8). Stress is the mismatch between the rank order of distances in the data, and the rank order of distances in the ordination. Samples grouped by source are oriented in such a manner that variance is minimized. SIMPER (Similarity percentage) breakdown was carried out to identify OTUs that contributed most to dissimilarity (Brays-Curtis distance) observed at 999 permutations.

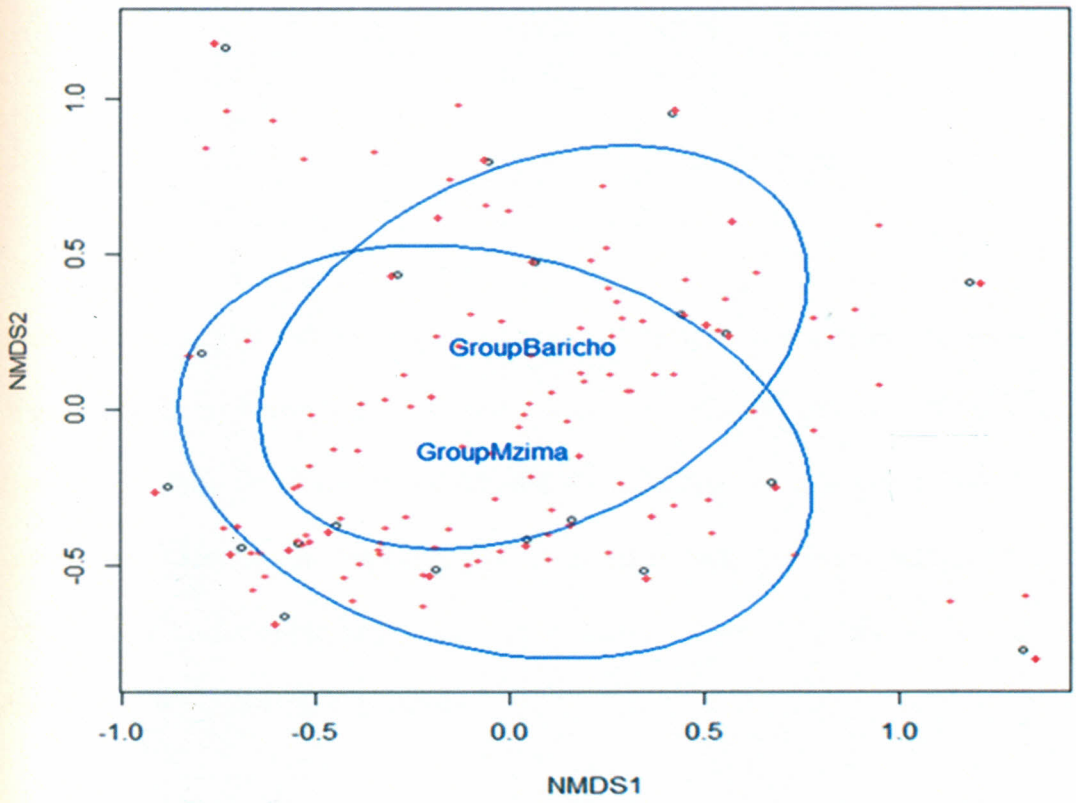


Fig. 4.8 Non metric dimensional scaling of bacterial class phylotypes clusters of Baricho and Mzima water lines

The highest levels of dissimilarity at class level was 82.5% between biofilms sampled from Baricho and Mzima water lines, followed by 82.01% between the water of Mzima water line and its biofilms (Table 4.5). Percentage of dissimilarity between Baricho water and Mzima line water was 80.49 and 78.55 between Baricho line water and its biofilms.

Table 4.5 SIMPER class dissimilarity using Brays Curtis Dissimilaity distances measure

	MZW	MZB	BRW	BRB
MZW		82.01	80.49	
MZB				82.5
BRW				78.55
BRB				

Redundancy analysis indicate that the classes of bacteria that contributed significantly to the dissimilarity between the two water sources (p= 0.019) and their different sample types (p= 0.026) are summarized (Table 4.6). Notable is the Candidate division's contribution to variation between the water lines both in bulk water, where they constituted all dissimilar phyla, and in biofilms where 3 of the 7 phyla causing disimillarity were candidate divisions.

Table 4.6 Redudancy analysis indicating the contribution of different classes to the the dissimilarity between different source and sample types

	Contribution	P
Mzima water and Baricho water		
Armatimonadetes	0.001010	0.014 **
Candidate division OP3	0.001010	0.014 **
Candidate division BRC1	0.000505	0.014 **
Candidate division WS3	0.000505	0.014 **
Mzima water and Baricho biofilms		
Proteobacteria	0.310866	0.080'
Candidate division TM7	0.014697	0.028 *
Candidate division OD1	0.010279	0.029 *
Elusimicrobia	0.008336	0.013 *
Nitrospirae	0.006377	0.015 *
NPL.UPA2	0.002262	0.063'
Candidate division OP11	0.002004	0.060'
SM2F11	0.001336	0.009 ***
TM6	0.000926	0.013 **
Chlorobi	0.000668	0.009 ***
Mzima biofilms and Baricho water		
Planctomycetes	0.028784	0.100'
Verrucomicrobia	0.012692	0.094'
Chloroflexi	0.010689	0.048 *
Acidobacteria	0.006526	0.078 '
Armatimonadetes	0.000970	0.084'
Candidate division OP3	0.000970	0.084'
Candidate division BRC1	0.000485	0.084'
Candidate division WS3	0.000485	0.084'
Mzima biofilms and Baricho biofilms		
Candidate division OD1	0.009725	0.066'
Elusimicrobia	0.007742	0.077'
Nitrospirae	0.006416	0.040 *
Cyanobacteria	0.004822	0.087'
SM2F11	0.001266	0.094'
TM6	0.000860	0.077'
Chlorobi	0.000633	0.094'

Signif. codes: $0 < p \leq 0.001$***, $0.001 < p \leq 0.01$**, $0.01 < p \leq 0.05$*, $0.05 < p \leq 0.1$'

Analysis of Similarities (ANOSIM) of the taxonomical profiles was used to establish whether the similarity values of the two water distribution systems based on bray Curtis distance measure at 999 permutations were significant. Evidently the profiles were not similar regardless of the taxonomic rank levels as all similarity p values were greater than 0.05 (Table 4.7).

Table 4.7 ANOSIM of the taxonomic profiles of Baricho and Mzima water distribution systems

Taxa	R	P
Class	0.08466	0.119
Order	0.06771	0.175
Family	0.05474	0.22
Genus	0.05053	0.197

NPMANOVA analyses testing the hypothesis of no difference between bacterial communities based on source type were rejected ($p = 0.028$). A 2 way NPMANOVA indicated significant interaction between bacterial community diversity and source type at 5% significance.

A heat map showing pairwise distance between all the sampling points relative abundances of phyla compared to those in other samples and the clustering of the samples based on similarity was generated. It also shows the phylogenetic clustering of the phyla using an embedded ultrameric phylogenetic tree. Four clusters of samples were generated based on the the relative abundance of phyla in the samples (Fig. 4.9). The first cluster was made up of MZB1, MZB2, MZB7, BRW4 and BRW6, predominated by *Proteobacteria* and *Firmicutes*. MZW4, MZW5, MZW6 and MZB4

were aligned into the second cluster where besides the *Proteobacteria* and *Firmicutes*, *Bacteroidetes* also show highest abudance. The BRB1, BRB2, BRW5 and MZW7 cluster has all the other predominant phyla and also *Actinobateria, Acidobacteria* and 2 candidate divisions (001 and TM7) being highly abundant. The most heterogenous cluster was made up of MZW1, MZW2, MZW3, MZB5, BRW1, BRB3 and BRB4 with high percentage of abundance of *Proteobacteria* and *Firmicutes* as indicated by high color intensity in each cell.

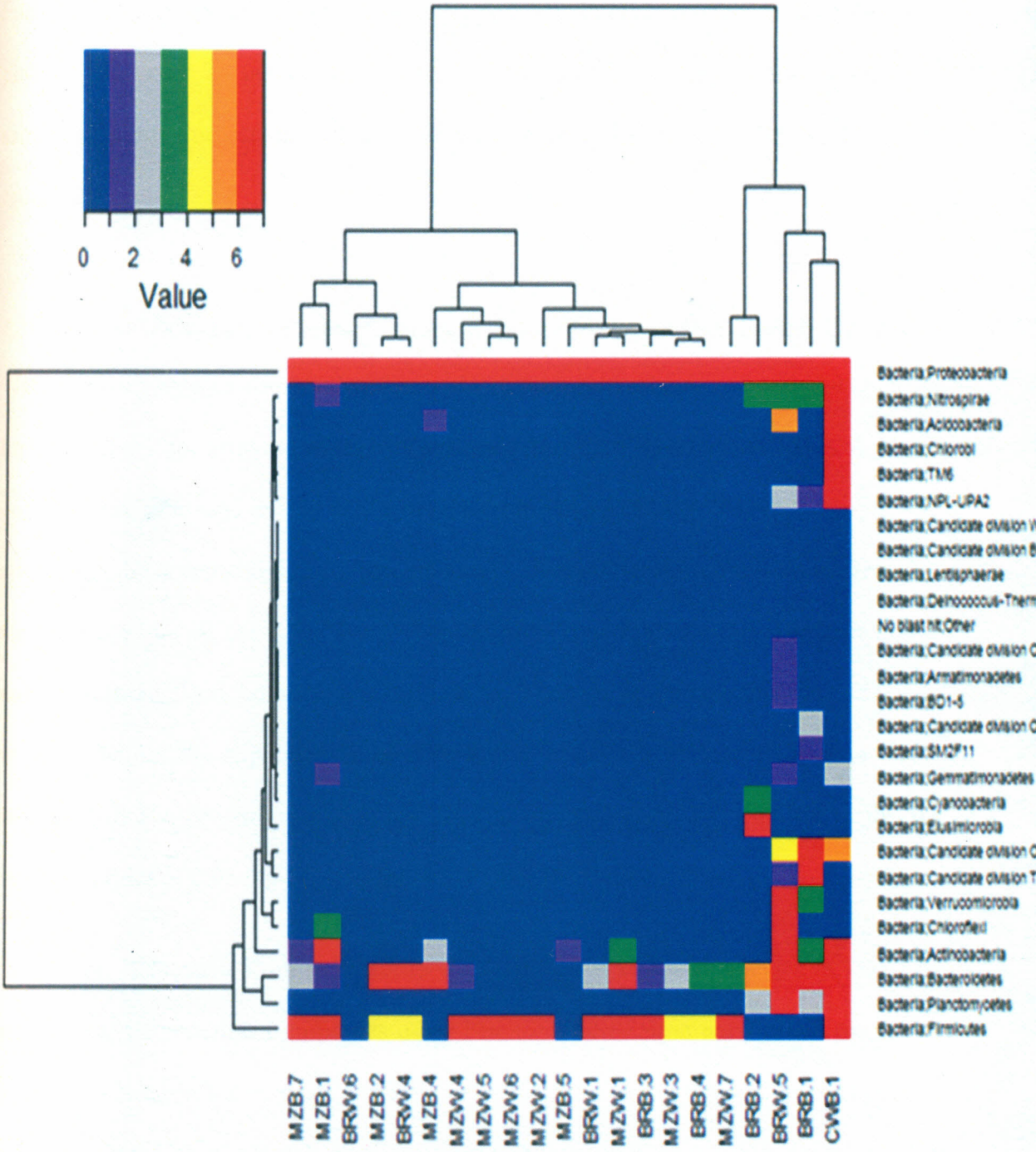


Fig. 4.9 Heat map showing relative abundance and hierarchical clustering of samples based on SILVA repository at 97% similarity of 16S rRNA gene V3 hypervariable region in Biodiversity R.

Each column represents a different sampling point while each row represents a different phylum; the color of the cells represents the frequency of that phylum within the sampling points (percentage of the sampling points from which the phylum was retrieved).

4.3 Physico-Chemical Characteristics of Water in Water Distribution Systems

Water from all sources complied with the pH, nitrates, phosphates and residual chlorine limits stipulated for drinking water by the Kenya Bureau of Standards (KEBS). A single occurrence of zinc was detected in a sample from BRB2 (Table 4.8). However, it was also within the acceptable limits. The maximum allowable iron in drinking water is 0.3 mg L^{-1} and while all the water samples from the Mzima line met this requirement, all samples from the Baricho water line had iron levels above acceptable limits. Lead levels were not within the stipulated acceptable limits in the 3rd, 4th and 7th Mzima water samples. All Baricho line samples had lead levels of 0.3 mg L^{-1} way above the set maximum limit of 0.05 mg L^{-1}, making water from the Baricho water line not potable.

Correlation tests carried out to establish the relationship between changes in levels of selected physico-chemical properties using Pearson's correlation gave varied results. A positive correlation between pH and Mn ($r = 0.942$, $p = 0.000$); temperature and pH, Fe and Mn ($r = 0.597$, $p = 0.004$; $r = 0.896$, $p = 0.000$; $r = 0.835$, $p = 0.000$ respectively) was observed. A negative correlation between temperature and NO_3–N ($r = -0.908$, $p = 0.000$); PO_4-P and Mn ($r = -0.440$, $p = 0.046$) nitrate nitrogen and pH, iron and

manganese ($r = -0.705$, $p = 0.000$; $r = -0.879$, $p = 0.000$ and $r = -0.821$, $p = 0.000$ respectively) was observed.

Table 4.8 Mean of physico-chemical characteristics of in the Mzima and Baricho water distribution systems of Mombasa City

Sample ID	Temp	pH	Residual Chlorine	NO_3-N	PO_4-P	Pb	Fe	Zn	Mn
MZW.1	23.3	7.65	0	4.3	0.01	0.025	0.006	ND	0.013
MZB.1	23.3	7.65	0	4.3	0.01	0.026	0.006	ND	0.013
MZW.2	23.6	7.74	0.08	3.91	<0.01	0.045	0.01	ND	0.014
MZB.2	23.6	7.74	0.08	3.91	<0.01	0.045	0.01	ND	0.014
MZW.3	24.2	8.02	<0.01	3.78	0.01	0.051*	0.02	ND	0.018
MZW.4	24.2	8.06	<0.01	2.94	0.04	0.05*	0.109	ND	0.02
MZB.4	24.2	8.06	<0.01	2.94	0.04	0.05*	0.17	ND	0.02
MZW.5	24.3	7.95	0.02	2.6	0.04	0.031	0.2	ND	0.02
MZB.5	24.3	7.95	0.02	2.6	0.04	0.031	0.2	ND	0.018
MZW.6	24.1	7.98	0.06	2.45	0.03	0.04	0.23	ND	0.02
MZW.7	25.2	7.99	0.14	2.41	0.06	0.05	0.14	ND	0.013
MZB.7	25.2	7.99	0.14	2.41	0.06	0.05	0.14	ND	0.013
BRW.1	26	8.12	0	1.28	0.01	0.303	0.599	ND	0.192
BRB.1	26	8.12	0	1.28	0.01	0.303	0.599	ND	0.192
BRB.2	26	8.09	0.19	1.41	0.02	0.377	0.8	0.085	0.204
BRW.5	25.6	7.84	0.11	1.52	0.01	0.315	0.67	ND	0.19
BRW.3	26.1	7.95	0.22	1.66	0.01	0.3	0.75	ND	0.176
BRB.3	26.1	7.95	0.22	1.66	0.01	0.3	0.75	ND	0.176
BRW.4	25.8	7.93	0.02	2.17	0.03	0.36	0.63	ND	0.11
BRB.4	25.8	7.93	0.02	2.17	0.03	0.36	0.63	ND	0.11
BRW.6	26.3	7.97	0.01	2.16	0.03	0.34	0.45	ND	0.102

Water temperatures ranged from 23.3°C to 26.3°C, with relatively higher temperatures in the Baricho water samples (26°C) as compared to Mzima waters (24°C). Median pH values in water samples from both water lines were almost similar with values of 7.9 and 7.8 in Baricho and Mzima respectively. Residual chlorine levels were quite variable ranging from no chlorine in some sources to 0.2 and 0.1 mg L^{-1} in Baricho and Mzima respectively.

Nitrate nitrogen levels were comparatively higher in Mzima water samples an unsually high value of 4.3 mg L^{-1} while the mean nitrate nitrogen values of Baricho samples was 1.7 mg L^{-1}. Phosphate phosphorus values were comparatively low at 0.01 and 0.03 mg L^{-1} in Baricho and Mzima respectively (Table 4.9). Lead, iron and manganese levels were abnormally high in Baricho yielding means of 0.33, 0.65 and 0.16 mg L^{-1} respectively, all of which are above acceptable limits of potability, while all Mzima water samples met the potablity requirements on lead, iron and manganese.

Table 4. 9 Physico-chemical characteristics of Mzima and Baricho water distribution systems

	Temp	**pH**	**Residual chlorine**	**NO_3-N**	**PO_4-P**	**Pb**	**Fe**	**Zn**	**Mn**
BWW	25.97± 0.206	7.99± 0.098	0.11±0.097	1.70± 0.375	0.02± 0.010	0.33± 0.031	0.65± 0.105	*0.09± 0.000	0.16± 0.042
Max	26.3	8.12	0.22	2.17	0.03	0.377	0.8	0.085	0.204
Min	25.6	7.84	0	1.28	0.01	0.3	0.45	0	0.102
MWL	24.13± 0.625	7.90± 0.156	0.08± 0.050	3.21± 0.764	0.03± 0.019	0.04± 0.010	0.10± 0.088	ND	0.02± 0.003
Max	25.2	8.06	0.14	4.3	0.06	0.051	0.23	0	0.02
Min	23.3	7.65	0	2.41	0.01	0.025	0.006	0	0.013

*outlier (single occurrence)

A comparison of the mean values of the physico-chemical properties of the different sampling sites of the two water lines using a one-way ANOVA test gave varied findings. There were no significant difference in the median values of pH, residual chlorine, phosphates and lead within and between the sampling points. However, significant differences were noted in the case of temperature, nitrates, iron and manganese ($df = 21$, $p<0.001$) between Mzima and Baricho water lines but not within individual water line samples.

Principal Components Analysis (PCA) resulted in projection of samples covariance matrix onto a set of axes using the Euclidean distance metric. The environmental variables representing the original axes are projected as biplot arrows which greatly aid in interpretation. Two distinct groups are clustered by water source type are evident. Mzima and Baricho water line are different in terms of their physic chemical characteristics. Phosphates were the critical determinant in the ordination of MZW7 water and biofilm (Fig. 4.10). Nitrates were the critical determinant in Mzima water at the source (MZW1 and 2), while temperature was key in BRW 6. Notably pH does not play a key role in the ordination of any of the samples, which implies that change in pH is not likely to change the taxa abundance at any sampling site.

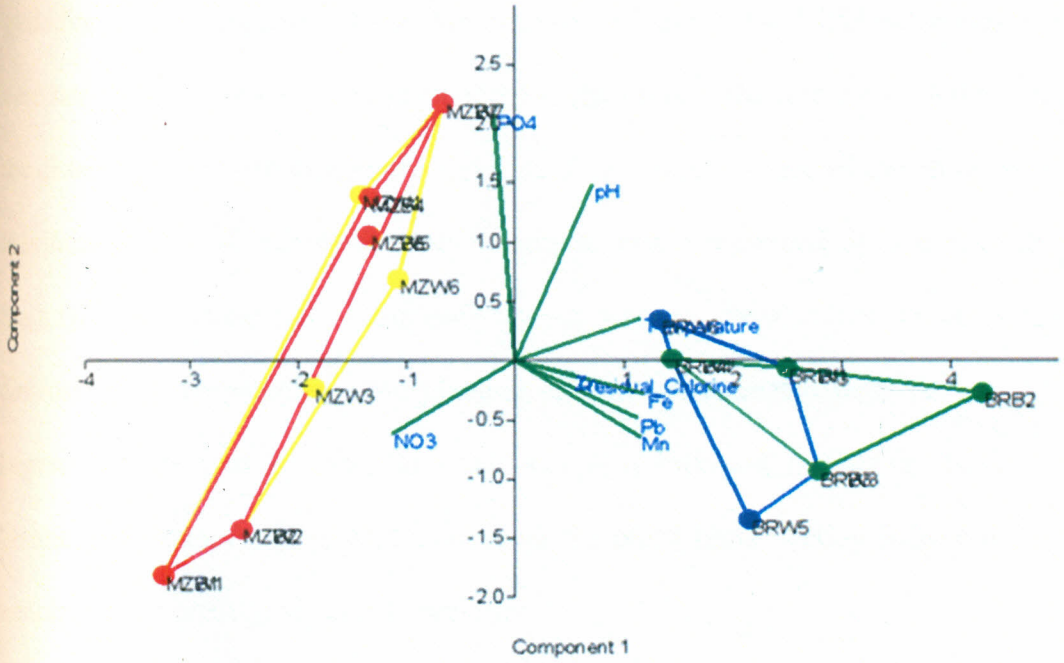


Fig. 4.10 Principal Component Analysis plots showing clustering of samples along critical environmental determinants of diversity

Red and yellow symbols correspond to biofilms and water from Mzima respectively. While blue and green represent Baricho line water and biofilms. Environmental variables were represented by arrows. The concentration of the parameters is continuously increasing in the direction of the arrow, and decreasing in the opposite direction.

Principal coordinates analysis (PCoA) was conducted to evaluate similarities of different samples using SILVA Classifier taxa, which were ordinated using their real distances rather than ranks with Bray Curtis distance measure. PCoA arranges samples between endpoints or 'poles' according to the distance matrix in an attempt to represent

the distances between samples. The PCoA algorithm is analogous to rotating the multidimensional object such that the distances (lines) in the 'shadow' are maximally correlated with the distances (connections) in the object. The linear correlation between the distances in the distance matrix, and the distances in a space of low dimension was maximized. The distance between two samples was represented by one dimension (a line). Distances between the samples in a group were represented in a similar number of dimensions as the samples. Four planular clusters were created from the PCoA, with largest distances among Baricho water samples, followed by Mzima biofilms and Mzima water samples (Fig. 4.11). However the plane hulls overlap indicating that the samples have very high levels of similarity.

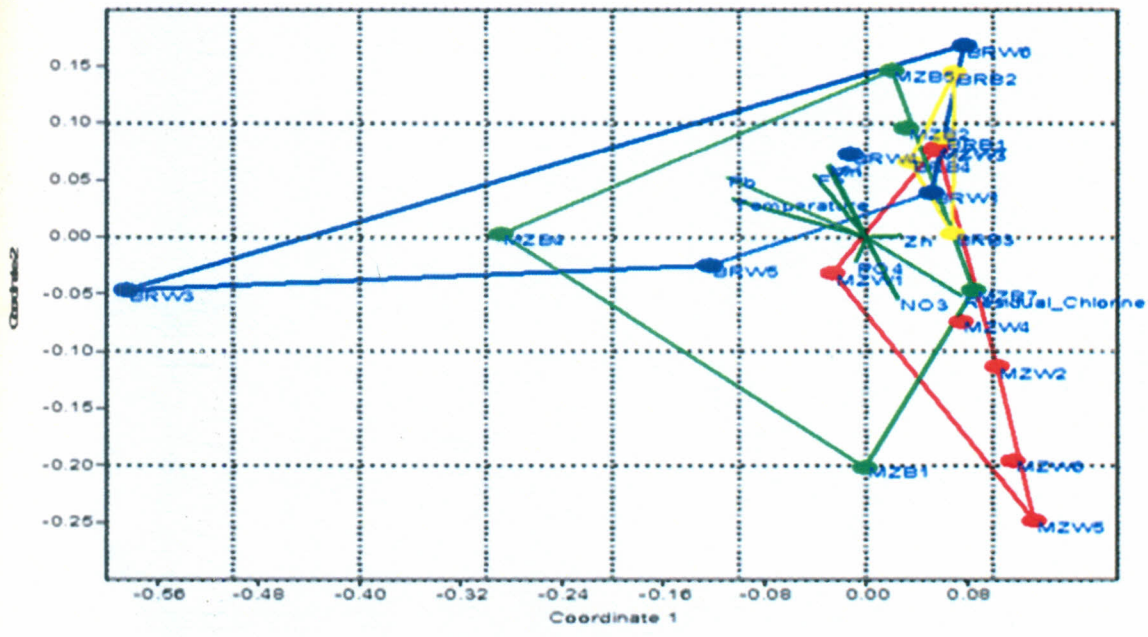


Fig. 4.11 Principal Coordinates Analysis showing similarity distances between sample types of Mzima and Baricho water lines.

Key: Red and yellow symbols correspond to biofilms and water from Mzima respectively. While blue and green represent Baricho line water and biofilms.

To discern the possible relationship between microbial community structure and environmental parameters, Canonical Correspondence Analysis (CCA) was performed (Fig. 4.12). Based on variance inflation factors with 999 Monte Carlo permutations, nine environmental variables, pH, temperature, residual chlorine, nitrate nitrogen (NO_3-N), phosphate phosphorus (PO_4-P), iron (Fe), manganese (Mn), Zinc (Zn) and lead (Pb), were selected in the CCA biplot. The length of an environmental parameter arrow in the ordination plot indicates the strength of the relationship of that parameter to community composition. As such, temperature, iron and nitrates appear to be the most important environmental properties in determining the composition and diversity of bacteria in all samples ($p<0.01$).

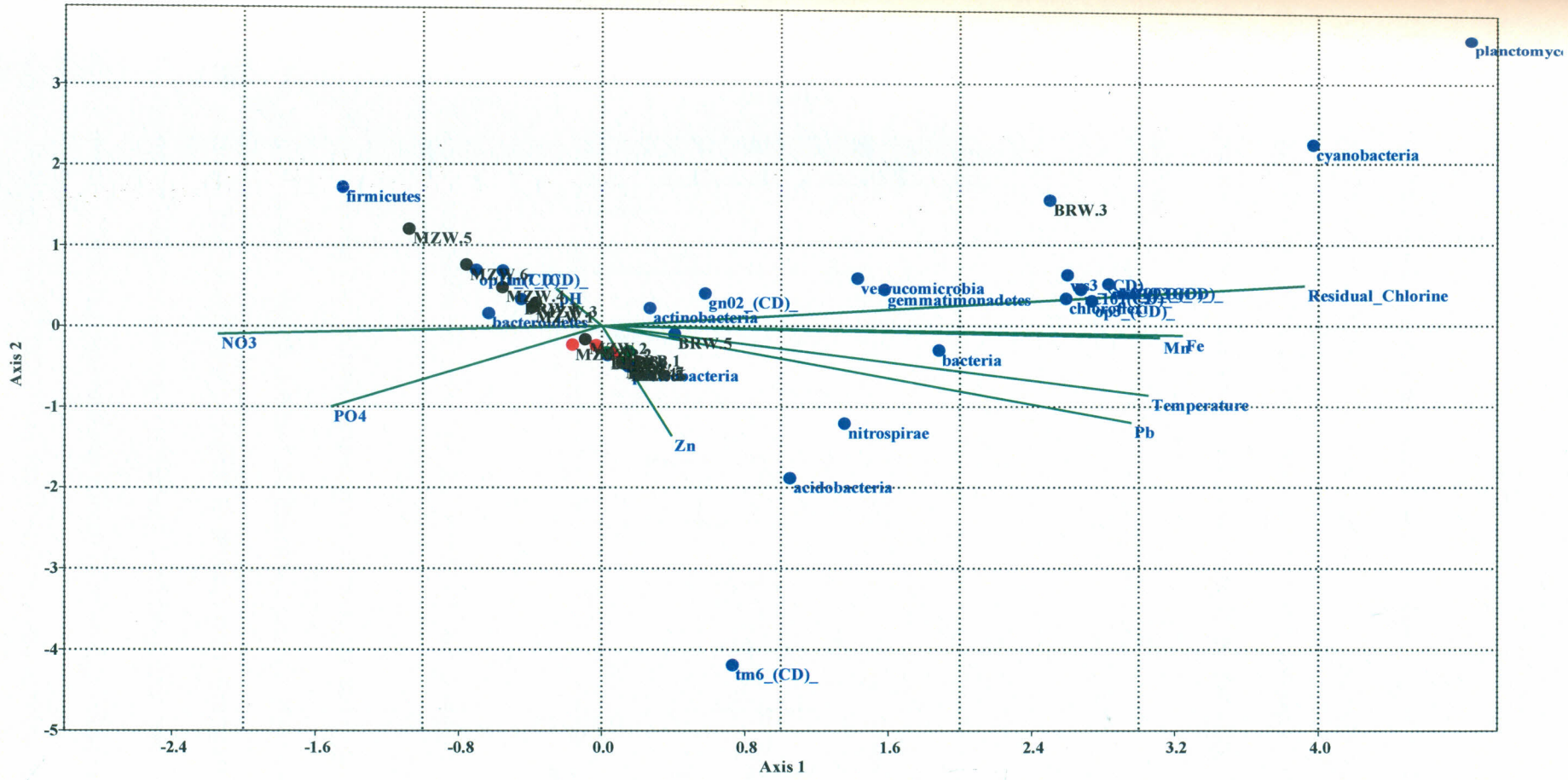


Fig. 4.12 Triplot from a CCA of the microbial community structure clustered along the physico-chemical variables influencing compostion and diversity

Environmental variables are represented by green arrows, samples (quadrats) by small red, yellow, blue and green circles. Arrows indicate the direction and magnitude of measurable variables associated with bacterial community structure.

4.4 Biofilm and Bulk Water Bacterial Communities

A comparison of bacterial composition between biofilms and bulk water from the two pipelines (Fig. 4.13) shows a clear difference as indicated by higher species richness in biofilms recovered from all sampling sites comparable to the cumulative richness in bulk water samples.

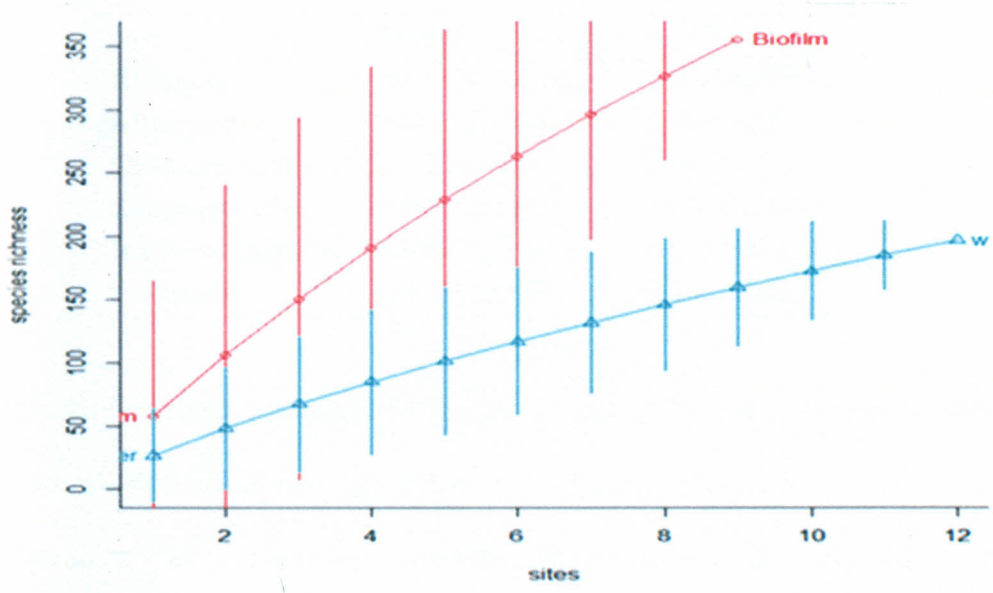


Fig. 4.13 Cumulative species richness of pooled water and biofilm samples

Comparison of bacterial community composition of biofilms and bulk water was carried out using Mantel statistic, which is based on Pearson's product-moment correlation with 999 permutations. The Mantel test was used to test the null hypotheses that the there is no relationship between sample type (water and biofilms), and bacterial diversity. The Mantle test revealed the existence of a significant relationship between sample type and diversity ($r= 0.412$, $p= 0.002$) indicating that bacterial communities of biofilm and bulk

water are separated in terms of dissimilarity but overlap due to some shared phylotypes. Hence the null hypothesis was rejected. Biofilms microbiomes had higher bacterial diversity richness, abundance, Shannon, Simpson and Inverse Simpson indices than water microbiomes, which on the other hand had a slightly higher evenness than biofilms (Table 4.10).

Table 4.10 Mean ±SD of diversity measures in water and biofilm samples at genus level

	Water	**Biofilm**
Richness	89.727±122.082	125.888±79.722
Abundance	552.272±554.713	620±43.58
Shannon index	3.326±0.667	3.632±0.320
Simpson index	0.9362±0.0195	0.940±0.0266
Inverse Simpson	18.902±12,442	19.383±6.801
Evenness	0.470±0.0970	0.363±0.072

Non Metric Dimensional Scaling (NMDS) plots were ordinated in two-dimensions to visualize overall similarities and differences in community structure between water and biofilm samples, by calculating pairwise Bray-Curtis and weighted UniFrac dissimilarities (Fig. 4.14).

Non metric multi dimension scaling ordination of class phylotype matrix yielded an acceptable representation of the observed distances of bacteria classes in bulk water and biofilms (final stress = 0.2011934) (Fig. 4.14). Several points are sparsely ordinated indicating that they are not similar ($p=0.013$). However extensive overlaps occur in the bacterial diversities of bulk water and biofilms.

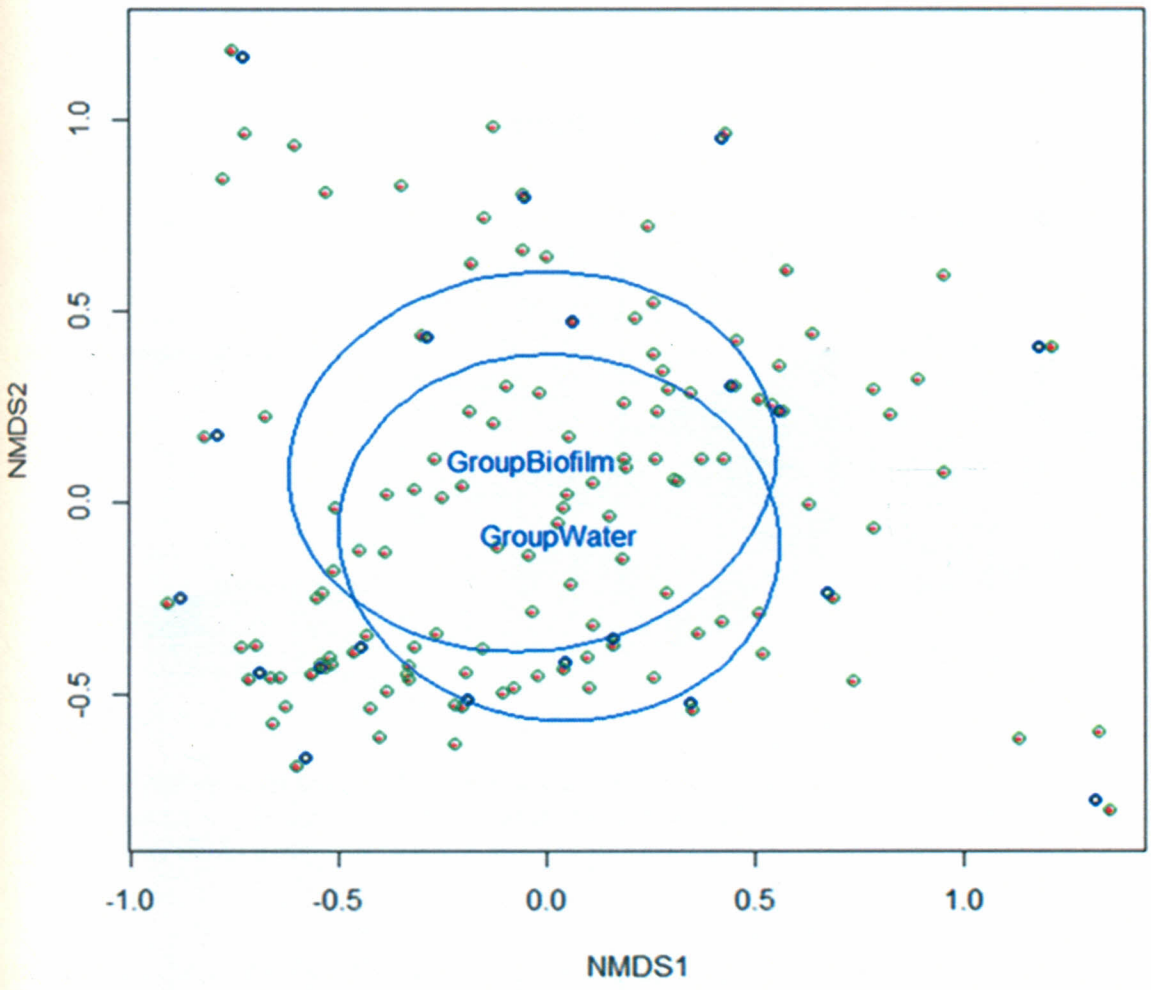


Fig. 4.14 NMDS plot of the two sample types illustrating their statistical compositional difference.

Comparing Baricho water and biofilm profiles at class level, minor differences were observed in the relative abundances of *Gammaproteobacteria* at 39.1% in water and 35.6% in biofilms. However, there were some notable differences in the relative abundances of *Nitrospira* (20.2%) in water and a mere 0.8% in biofilms (Fig. 4.15). Other dominant classes in Baricho water were *Betaproteobacteria* at 15.9% followed by *Alphaproteobacteria* at 6.4%, while *Alphaproteobacteria* (29.6%), *Betaproteobacteria* (11.1%) and *Deltaproteobacteria* (3.2%) were prominent in biofilms.

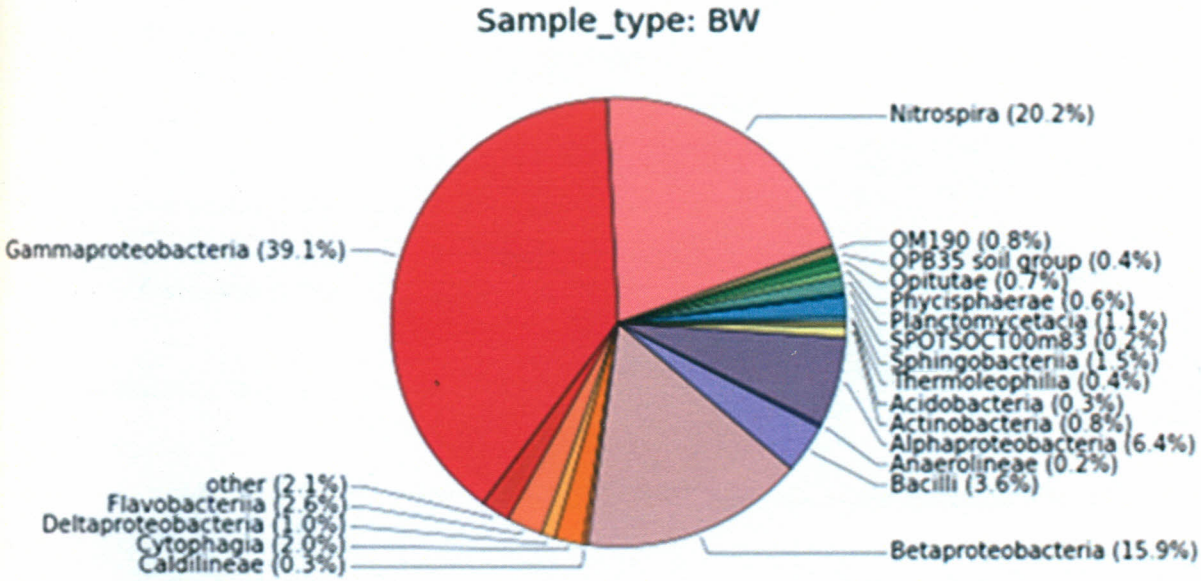


Fig. 4.15 Relative abundance of different bacteria classes in Baricho water to the total bacterial phylotypes based on Metageneassist analysis 454 pyrosequencing of the 16S rRNA gene.

Bacilli were the most abundant classes of *Firmicutes* in both microbiomes making up 6.4% and 3.6% of the total number of analyzed reads, in biofilms and water respectively (Fig. 4.16). The biofilm profile uniquely yielded the uncultured bacterial classes (3%).

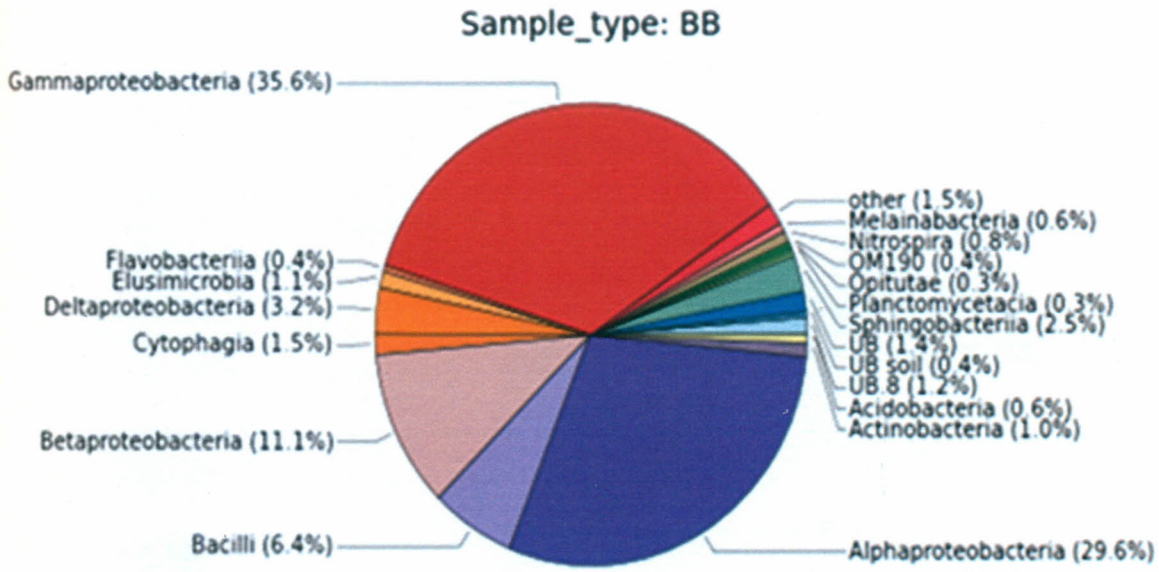


Fig. 4.16 Relative abundance of different bacteria classes in Baricho biofilms to the total bacterial phylotypes based on Metageneassist analysis 454 pyrosequencing of the 16S rRNA gene

The bacterial profiles diversity of biofilms and water sample of the Mzima water line were also compared. Comparing Mzima water and biofilm profiles at the class level, minor differences were observed in the relative abundances of *Gammaproteobacteria* (41.1% in water and 38.0% in biofilms), *Betaproteobacteria* (23.8% in water and 27.3% in biofilms) and *Alphaproteobacteria* (8.7% in biofilms and 4.3% in Mzima water). *Bacilli* and *Clostridia* were the most abundant classes of *Firmicutes* in both profiles making up 24.3% and 2.1%; and 6.4% and 4.7%, of the total number of analyzed reads in water and biofilms respectively (Fig. 4.17).

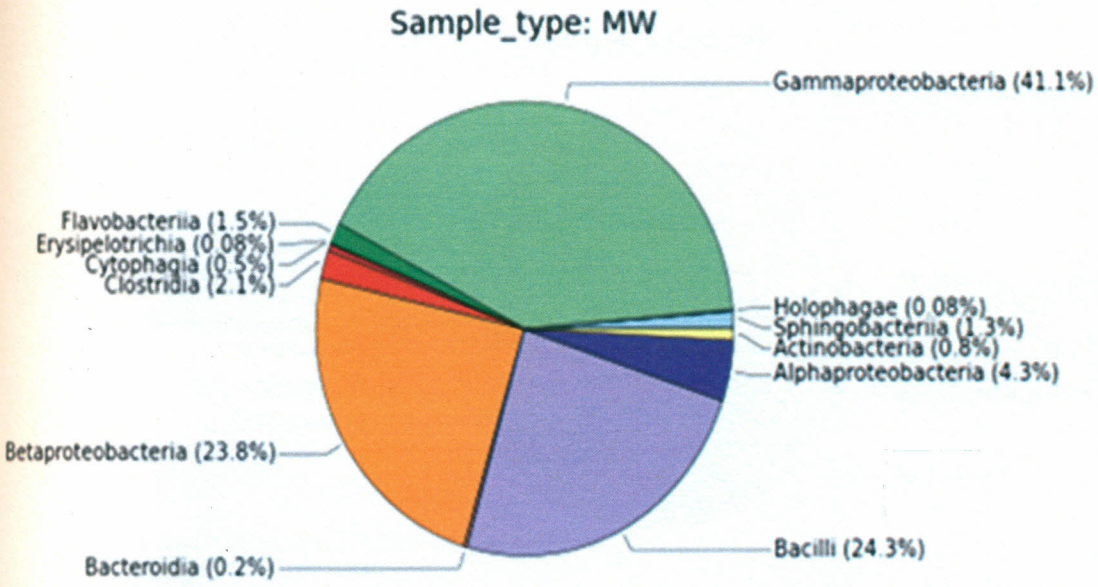


Fig. 4.17 Relative abundance of different bacteria classes in Mzima water to the total bacterial phylotypes based on Metageneassist analysis 454 pyrosequencing of the 16S rRNA gene

Biofilms featured high proportions of *Flavobacteria* (8.6%) and *Deltaproteobacteria* at 2.3% comparable to water profiles that had 1.5% *Flavobacteria* with no *Deltaproteobacteria* being identified in Mzima water samples (Fig. 4.18).

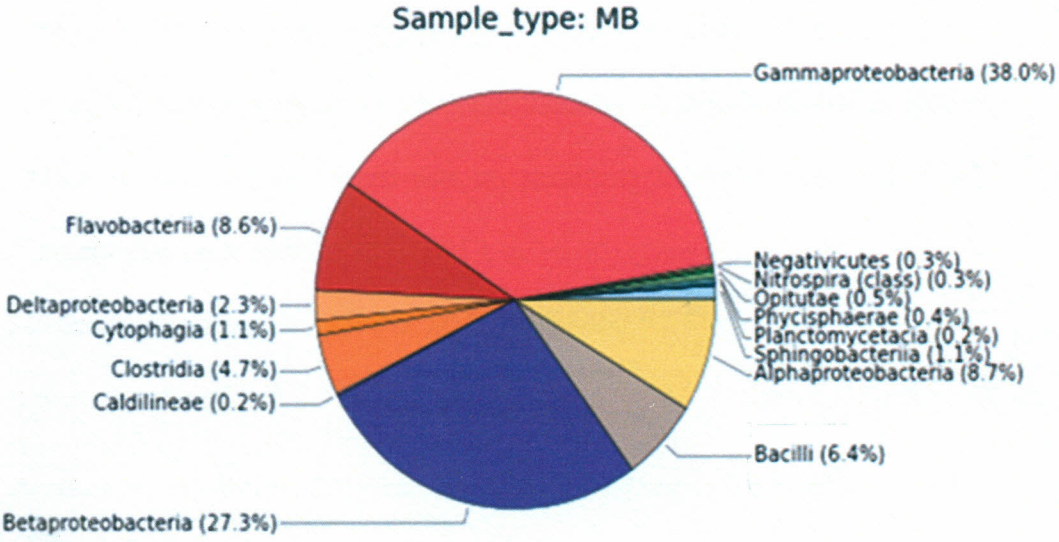


Fig. 4.18 Relative abundance of different bacteria classes in Mzima biofilms to the total bacterial phylotypes based on Metageneassist analysis 454 pyrosequencing of the 16S rRNA gene

Redundancy analyses (RDA) was computed to find out the bacteria phyla that contribute significantly to the dissimilarity between the bacterial profiles of the different sample types from the two water sources (p=0.026) (Table 4.11).

Table 4.11 Redudancy analysis indicating the contribution of different phyla to the the dissimilarity between different source and sample types

	Contribution	**p**
Mzima water and biofilms		
Firmicutes	0.090340	0.015 *
Actinobacteria	0.033299	0.052'
Acidobacteria	0.005986	0.071'
Baricho water and biofilms		
Candidate division OD1	0.009848	0.096'
NPL UPA2	0.002669	0.058'

Signif. codes: $0 < p \leq 0.001$***, $0.001 < p \leq 0.01$**, $0.01 < p \leq 0.05$*, $0.05 < p \leq 0.1$'

Analysis of similarities (ANOSIM) was used to check for levels of similarity between the bacterial profiles of the two sample types from the two water distribution systems based on Bray Curtis distance measure at 999 permutations. Analysis of similarity test at 999 permutations confirmed that there was no similarity between bulk water and biofilms at all taxonomic rank levels (R: -0.0919, p >0.05). Class similarity of 0.0899 was observed in the 90% quantile, 0.1300 at 95%, 0.1836 at 97.5%, while at 99% only 0.2411 classes were similar (Table 4.12).

Table 4.12 Analysis of similarities of the taxonomic profiles of respective biofilm and water samples of different taxonomic ranks

	R	P
Class	0.000999	0.392
Order	0.0131	0.492
Family	0.01709	0.534
Genus	-0.01576	0.526

Using a Non-Parametric multivariate analysis of variance (NPMANOVA) test to evaluate the difference in bacterial communities, the hypothesis that there is no difference between bacterial communities obtained from different sources was rejected (F= 0.8035, p = 0.8489). A two way NPMANOVA indicated that there were no significant interaction between bacterial community diversity and sample type. The levels of dissimilarity between bacterial classes in water and biofilms (Table 4.13) consistently increased from the lower to upper quartiles of the sequences representative of the samples.

Table 4.13 Dissimilarity ranks between and within classes based on sample type

	0%	25	50	75	100	N
Between	1	48	102	152	209	108
Biofilm	6	47.8	94	140	193	36
water	4	71.2	124	172	210	66

4.5 Patterns of Variation of Bacterial Communities in Different Sections of the Water distribution System

Comparative analysis of the microbiomes at the source and endpoint of Mzima and Baricho water line revealed that as the pipeline runs its course, some changes in phyla composition both in diversity and the proportion of OTUs occurred. *Proteobacteria* remain the dominant phyla at all levels, making up 81% and 68.6% of the bacterial composition in Baricho and Mzima water. *Firmicutes* contributed 13% and 12.9% while the *Bacteriodetes* contributed 5% and 12.3% in Baricho and Mzima water respectively (Fig. 4.19). At the Mzima water source, the proportion of *Actinobacteria* was 5.7%. The *Proteobacteria* also dominated the Baricho water end point where it accounted nearly all the OTUs (100%). A total of eight phyla, most of which occurred in biofilms, were present only at source but not at endpoints. These phyla comprised *Nitrospirae, Elusimicrobia, Cyanobacteria, Gemmatimonadetes, NP-UPA2,* and Candidate Divisions *TM7, OP11* and *OD1*. A two way ANOVA test confirmed the existence of a significant difference in bacterial phyla diversity ($p = 0.0111$) between source and end point biofilms but no significant difference between source and endpoint water phyla ($p=0.814$).

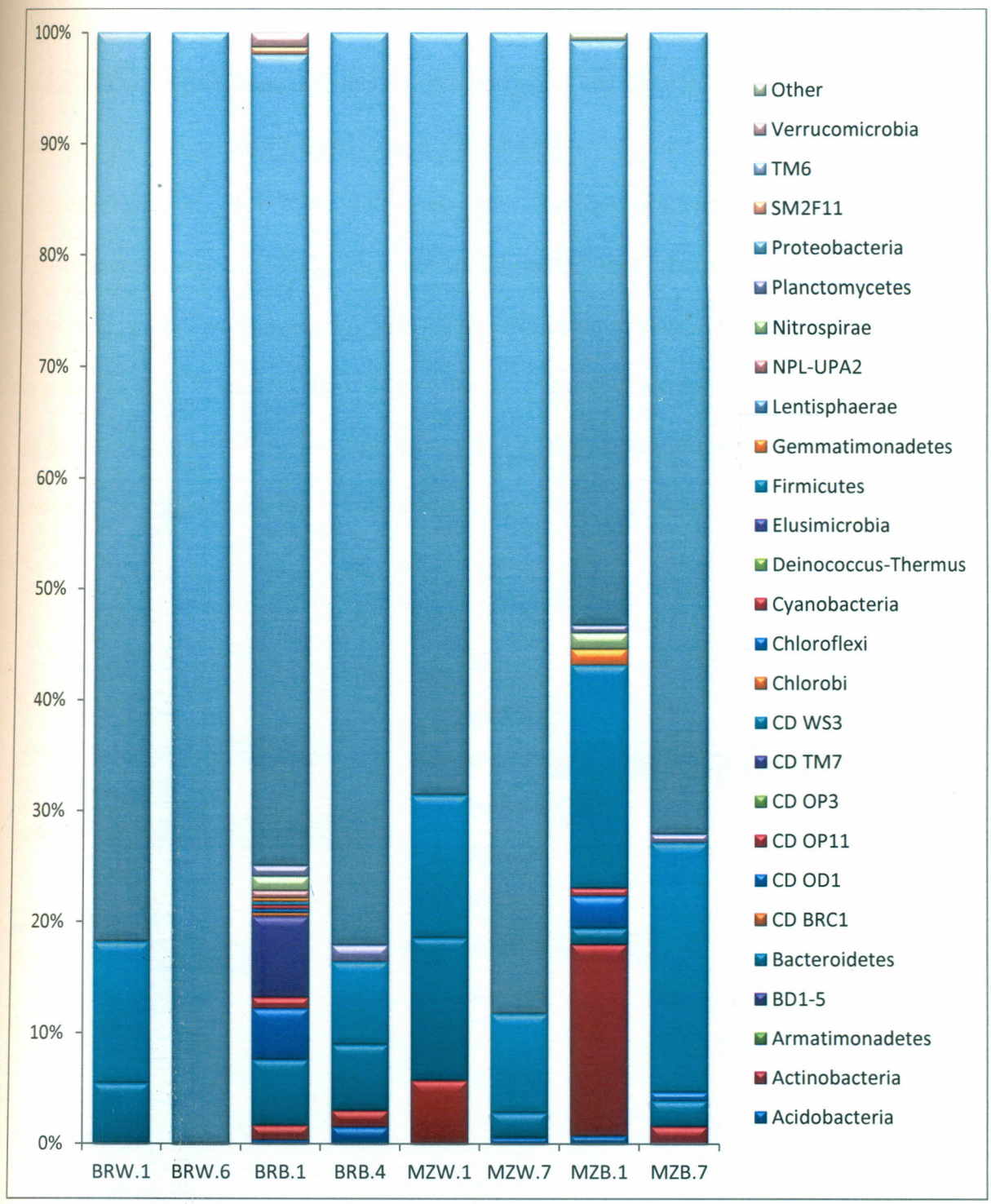


Fig. 4.19 Relative abundance of phyla at the source and endpoint of Baricho and Mzima water lines based on SILVA 119 reference database

The distribution of bacterial classes at the source and endpoint (household water meters) of the different water lines (Mzima and Baricho), and the different sample types (water and biofilms), gave varied results (Fig. 4.20). Mzima water at source was dominated by *Betaproteobacteria* (28.7%) followed by *Alphaprotobacteria* (24.3%) and *Gammaproteobacteria* (15%). The end point water was characterized by a high proportion of *Betaproteobacteria* (67%). The proportion of *Gammaproteobacteria* also increased to 21%. Present at the source water but missing at the endpoint water were the *Alphaproteobacteria*. On the other hand, *Clostridia*, which contributed 8.4% were present in end point water but not at source water.

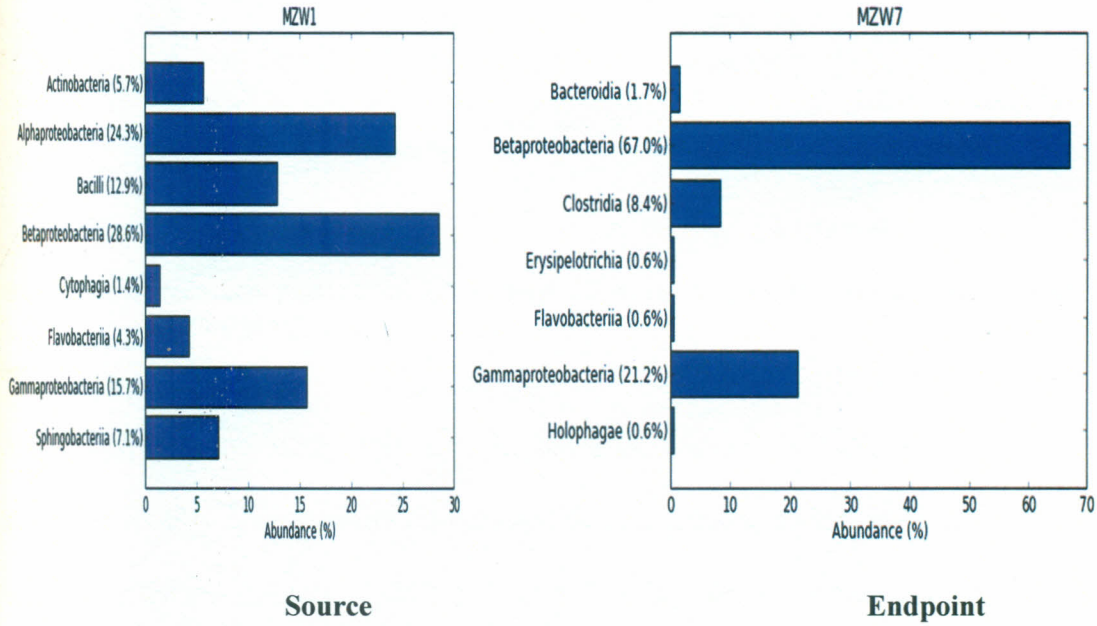
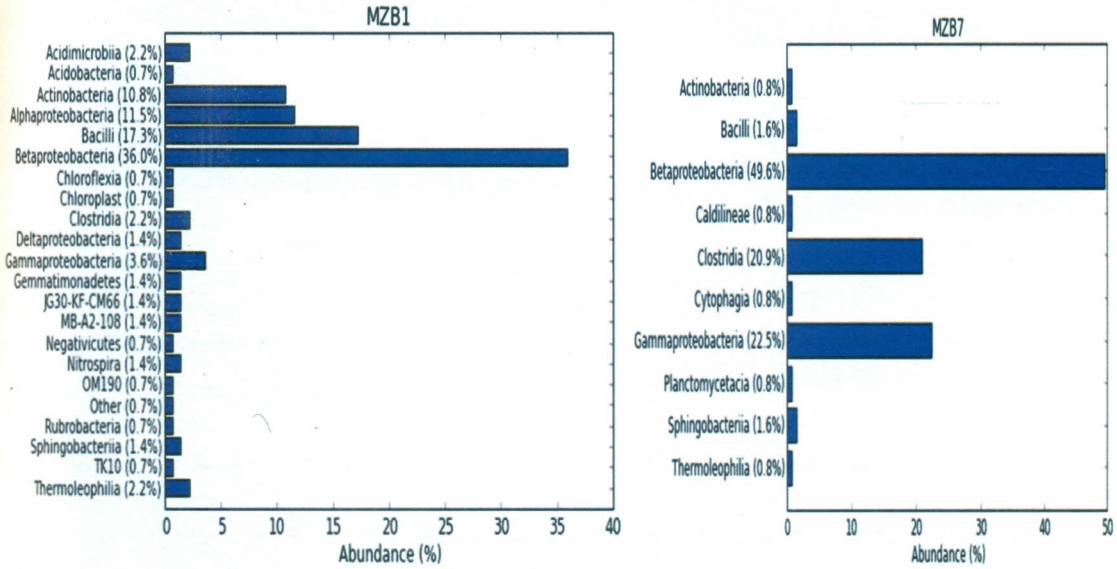


Fig. 4. 20 Percentage distribution of bacteria class phylotypes in Mzima bulk water at source and endpoint

Mzima biofilms were dominated by *Alpha, Beta* and *Gamma Proteobacteria* both at source and endpoint although the proportion of the *Gammaproteobacteria* increased

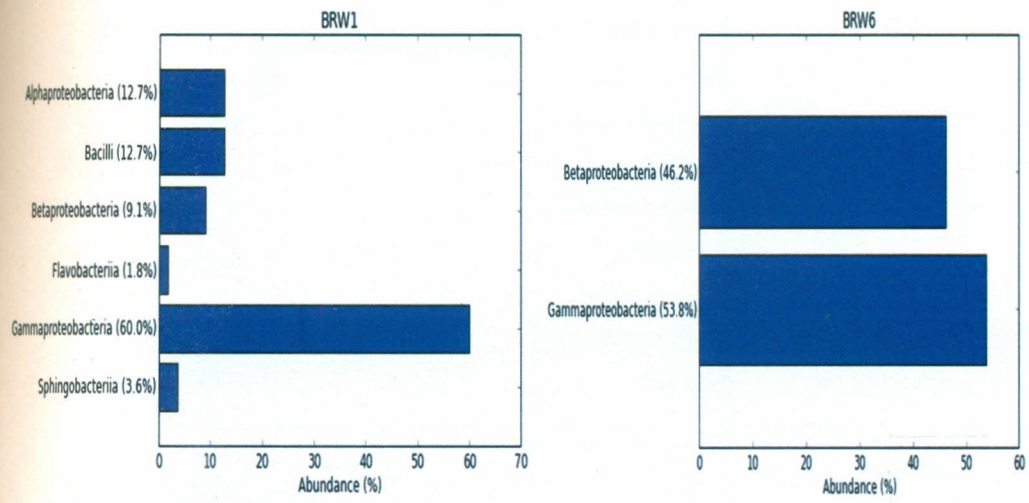
from 3.6% at the source to 22.5% at the endpoint (Fig. 4. 21). Also notable was the increase in class *Clostridia* from 2.2% in biofilms at source to 20.9% in endpoint biofilms. Source biofilms had a greater bacterial diversity with more than 20 classes represented while end point biofilms feature around 10 different bacterial classses.



Source **Endpoint**

Fig. 4. 21 Percentage distribution of bacteria class phylotypes in Mzima water source and endpoint biofilms

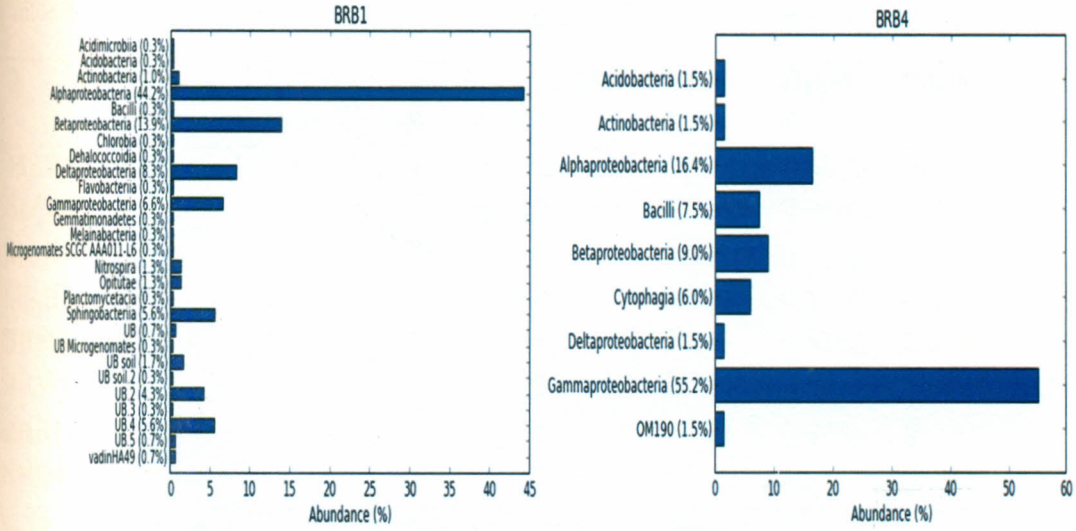
Baricho water was noticeably low in bacterial classes and endpoint water had a characteristically low diversity represented by two classes, *Betaproteobacteria* (46.2%) and *Gammaproteobacteria* (53.8%) (Fig. 4. 22). Source water was also dominated by *Gammaproteobacteria* (60%) while *Betaproteobacteria* and *Alphaproteobacteria* contributed 9.1% and 12.7% of the phyotypes respectively. *Bacilli* (12.7%) was present only in source water but not the endpoint water.



Source **Endpoint**

Fig. 4. 22 Percentage distribution of bacteria class phylotypes in Baricho bulk water at source and endpoint.

Baricho biofilms revealed a greater bacteria class diversity at source, as compared to the end point although most of the classes had very low abundance of less than 1% of the phylotypes. The most abundant class was *Alphaproteobacteria* (44.2%) followed by *Betaproteobacteria* (13.9%) *Deltaproteobacteria* (8.3%) and *Gammaproteobacteria* (6.6%) (Fig.4.23). Also notable was the presence of 8 classes of the uncultured bacteria and *Sphingobacteria* at 5.6%. The endpoint biofilms were dominated by *Gammaproteobacteria* (55.2%) *Alphaproteobacteria* (16.4%) and *Betaproteobacteria* (9%). A notable increase in *Bacilli* from 0.3% at the source biofilms to 7.5 at end point biofilms.



Source **Endpoint**

Fig. 4. 23 Percentage distribution of bacteria class phylotypes in Baricho source and endpoint biofilms

Pairwise comparisons of water and biofilm samples from both Mzima and Baricho waterlines at source and at endpoint meters confirmed that the bacterial communities of most sampling sites where more or less similar (Table 4.14). Significant differences were noted in the source water and source biofilms of the Baricho line ($p= 0.013$) and between source and endpoint biofilm bacterial communities of the Baricho water line ($p= 0.025$).

Table 4.14 Wilcoxon pairwise comparisons of sample types at source and endpoint

0	MZW-S	MZW-E	MZB-S	MZB-E	BRW-S	BRW-E	BRB-S	BRB - E
MZW1	0	0.814	0.3538	0.8908	1	0.6923	0.2238	0.5814
MZW7	1	0	0.5198	0.7154	0.5692	0.4462	0.013*	0.9455
MZB1	1	1	0	1	0.3112	0.1259	0.3576	0.3229
MZB7	1	1	1	0	0.5331	0.1829	0.1727	0.1579
BRW1	1	1	1	1	0	0.8182	0.013*	0.8462
BRW6	1	1	1	1	1	0	0.013*	0.5814
BRB1	1	0.364	1	1	0.364	0.364	0	0.02032*
BRB4	1	1	1	1	1	1	0.5691	0

S- Waterline source **E**- Waterline end or userpoint

Nitrates and phosphates levels were higher in Mzima source water comparable to water at the endpoint. The levels of lead increased slightly in at endpoints of both Mzima and Baricho water comparable to the sources (Fig. 4.24). Iron and manganese levels were higher in the Baricho source water with a notable decline at the endpoint. The reverse was visible for Iron, which notably increased at the endpoint comparable to the source water in Mzima.

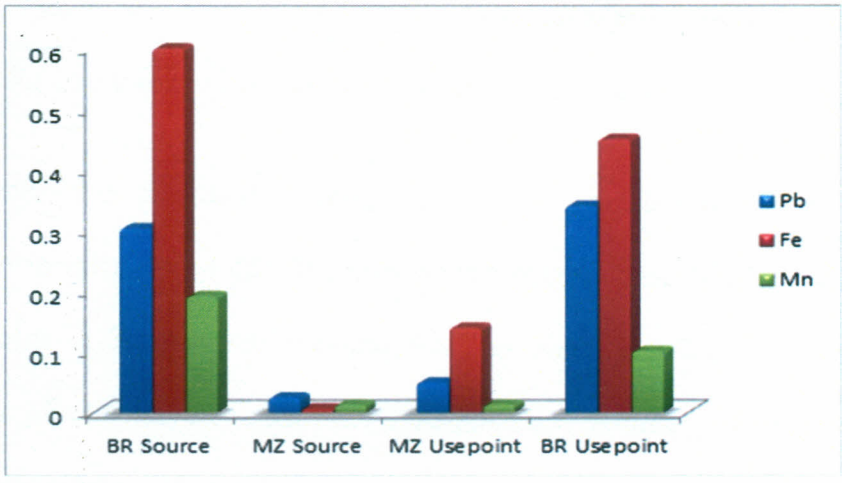


Fig. 4. 24 Levels of metal ions in Mzima and Baricho water lines at source and end point.

There was no significant difference in the physico-chemical characteristics of water at source and endpoint (p=0.0494). Residual chlorine levels were definitely higher at the endpoint due to chlorination before endpoint distribution, but no trace of chlorine in source water as a result of no chlorination.

4.6 Pathogenic Species in Water Distribution Systems

Blasting against the Green gene`s database at 97% sequence similarity threshold generated a total of 617 species from both the bulk water and biofilms of the two distribution systems. Although it is impossible to predict the pathogenicity of a strain based on 16S rRNA sequences, several sequences retrieved from the metagenomes were associated with bacterial groups with potential public health relevance. Some genera, such as *Acinetobacter Stenotrophomonas*, *Burkholderia, Flavobacterium, Pseudomonas, Staphylococcus* and *Escherichia/Shigella,* which are known for their facultative pathogenic and nosocomial character, were identified. Gram-negative pathogens, such as *E. coli, Klebsiella, Pseudomonas, Serratia, Enterobacter, Edwardsiella, Proteus* and *Chryseobacterium* were also identified.

Approximately 146 potentially pathogenic species were identified from the 21 water and biofilm samples (Fig. 4.25). The most abundant pathogens in different Mzima water line samples were *Bacillus thureingensis, Bacillus niacin, Bacillus flexus, Pseudomonas aeruginosa, Pseudomonas stastzeri, Pseudomonas putida, Pseudomonas guinea, Pseudomonas mosselii, Eenterobacter hormaechei, Bdellovibrio species, Hymenobacter species, Acinetobacter johnsonii, Aeromonas veronii, Aeromonas taiwanensis* and

Clostridium tunisiense. The dominant potentially pathogenic species on the Baricho water line were *Bacillus thureingensis, Pseudomonas stastzeri, Pseudomonas guinea, Aeromonas veronii, Aeromonas taiwanensis, Acinetobacter johnsonii, Candidate division TM7, Pseudomonas alkalygenes, Aeromonas hydrophila, Dermatophilus species, Staphylococcus scivius* and *Bacillus cereus.*

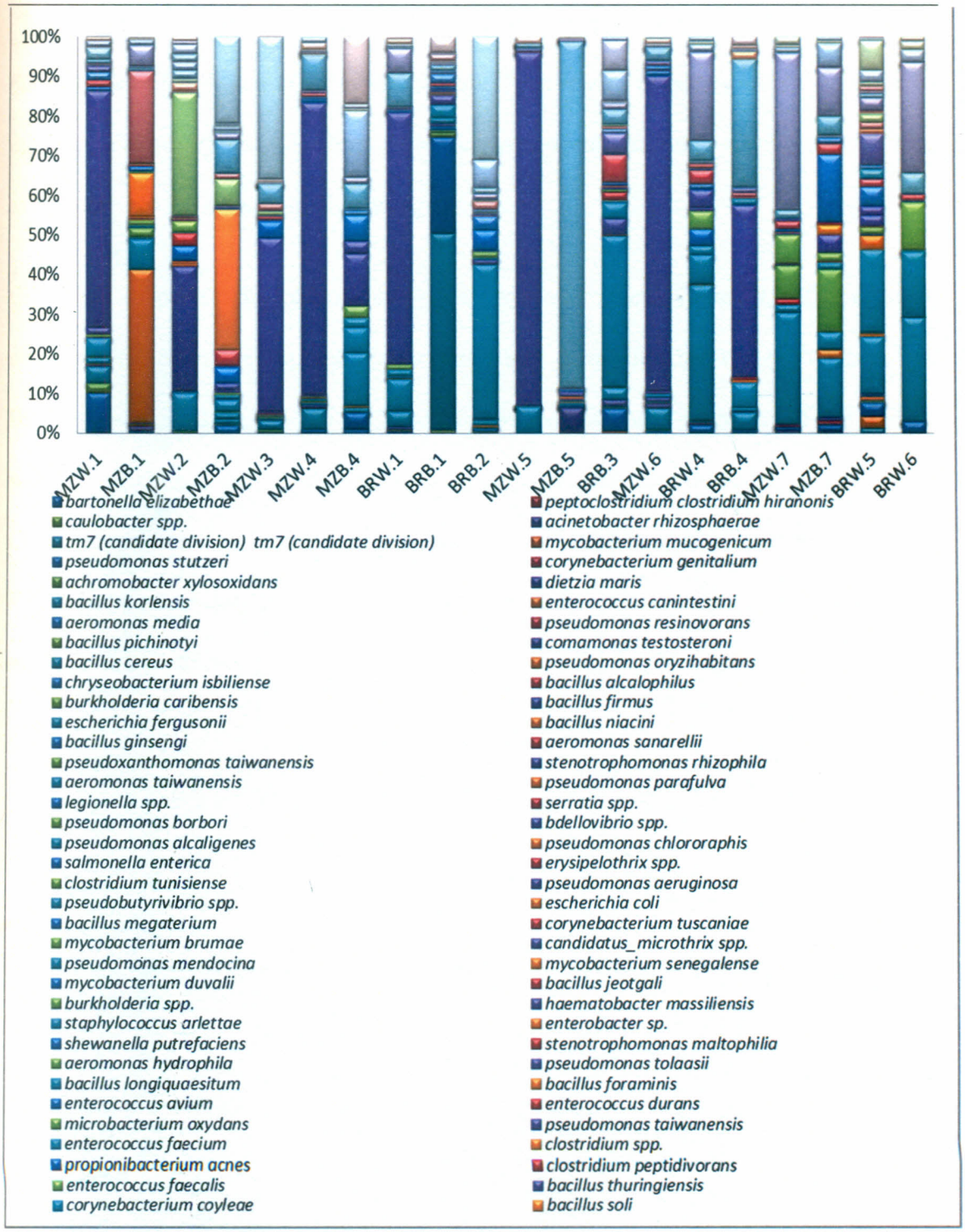


Fig. 4. 25 Relative abundance of pathogenic species of the water distribution

The metagenomic data indicates that *Mycobacterium* spp. occur in lower abundance in the water samples (found in BRW 4 and 5 only) but is very abundant in biofilms of both the Mzima and Baricho water lines. The truly pathogenic bacterial species are very few as compared to the opportunistic pathogens.

Ordination of potentially pathogenic species by water lines (Stress: 0.1782518) with obvious overlaps indicating that a large percentage of species occur in both Mzima and Baricho water lines (Fig. 4.26). *Mycobacterium* species notably ocurred in the Baricho water line while most of the *Bacillus* species featured in the Mzima water line. Clustering partterns clearly show ordination of simillar pathogens closer to one another. Regardless of the water line the *Bacillus* species and various species of *Mycobacterium* are ordinated in similar planes.

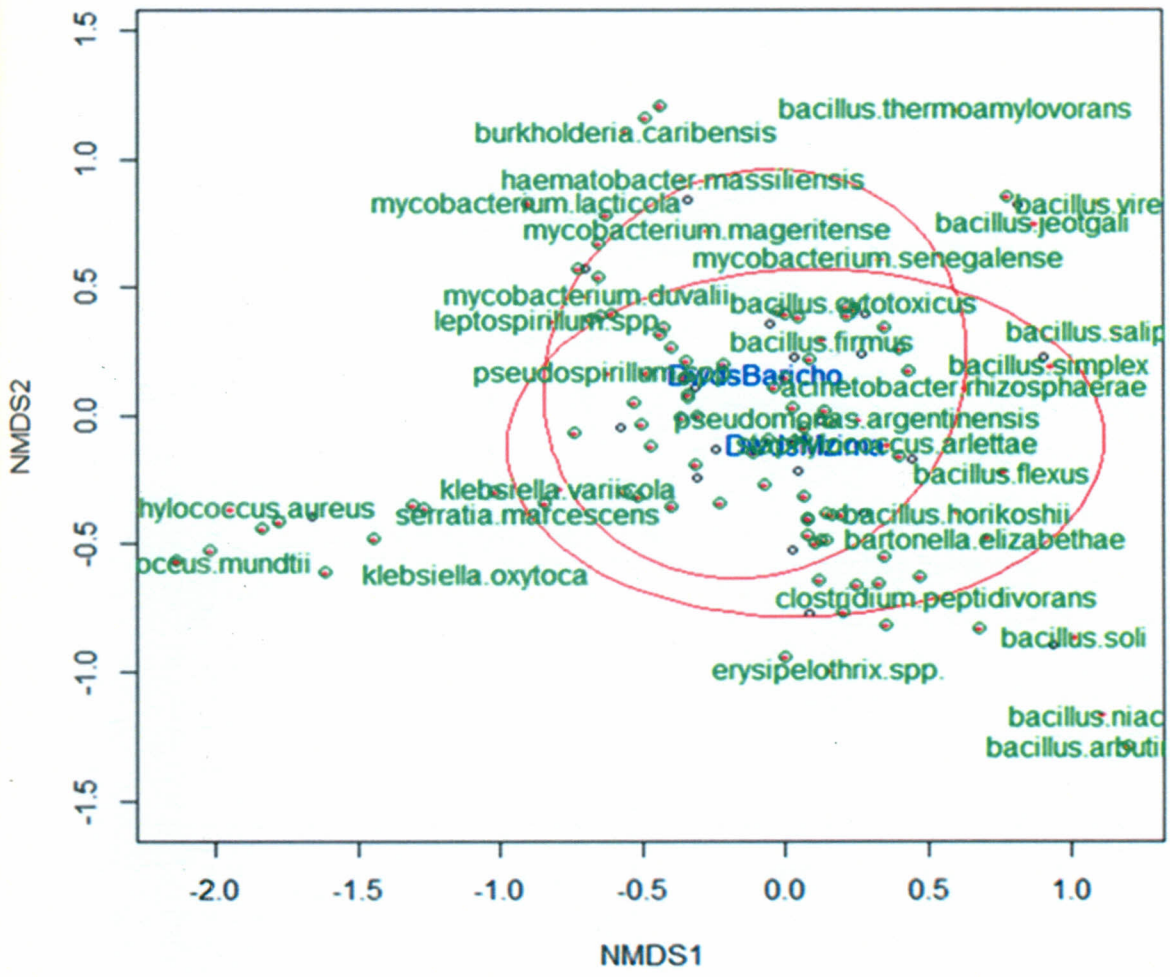


Fig. 4. 26 NMDS of pathogenic species in Mzima and Baricho water lines

Ordination of pathogenic species by sample type indicates a good representation of distance among the two water lines (Stress: 0.1782497**).** Most pathogenic species were detected in both water and biofilm samples (Fig. 4. 27). *Klebsiella, Leptospirillum* and *Serratia* species were unique to water samples while different species of *Bacillus* and *Clostridium* were unique to the biofilms. Again *Bacillus* species are ordinated along the right plane showing higher abundance in biofilm samples.

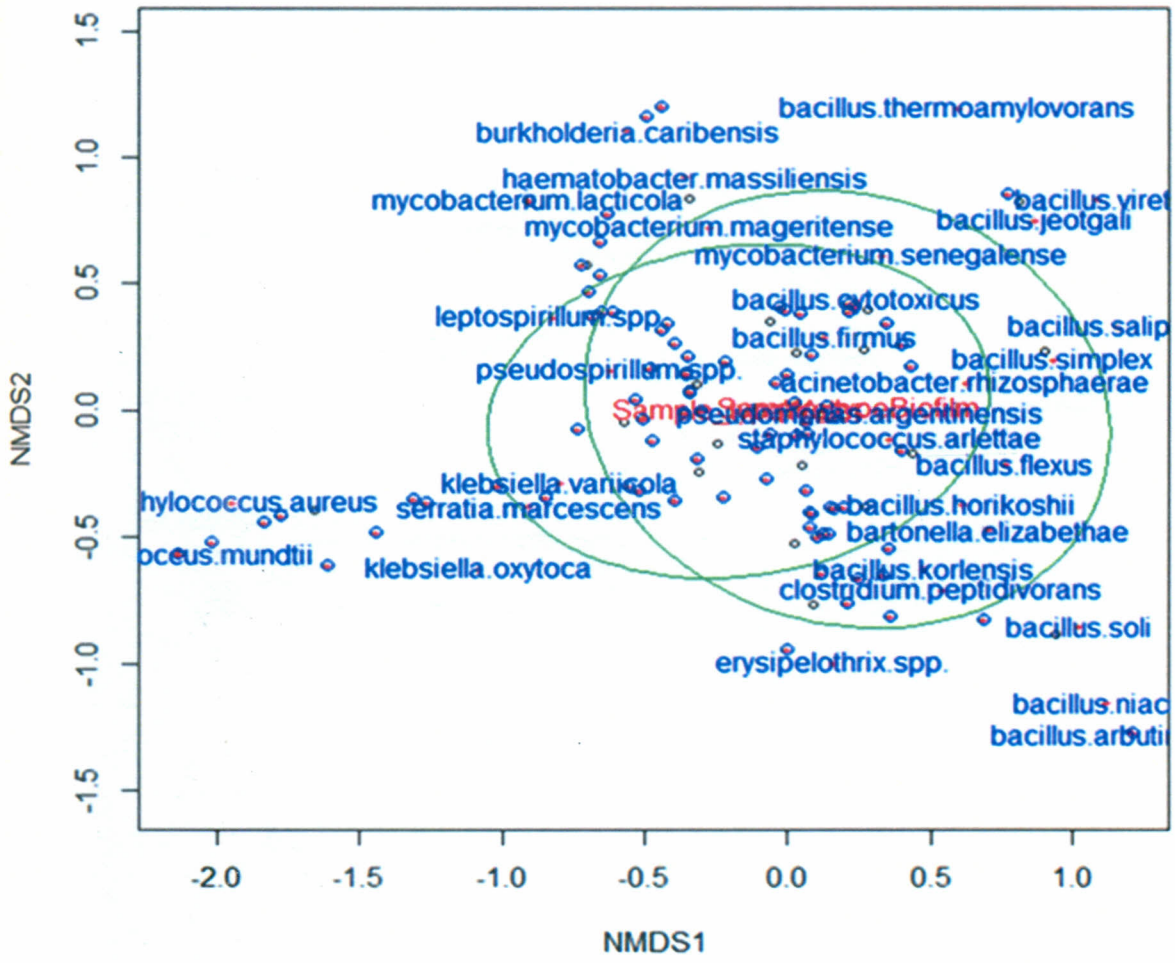


Fig. 4. 27 NMDS of pathogenic species in bulk water and biofilms

Both water lines had fairly similar species of pathogenic bacteria which is a good fit of the diversity by sample type (Stress: 0.1360182), with *Leptospirilla, Bdellovibrio* and *Corynebacteria* species being unique to the Baricho water line while *Klebsiella, Shigella, Aeromonas* and *Clostridium* species were unique to Mzima water line (Fig. 4.28). Generally the pathogenic species are not ordinated close to each other and bulk water and biofilms pathogenic composition is therefore not simillar.

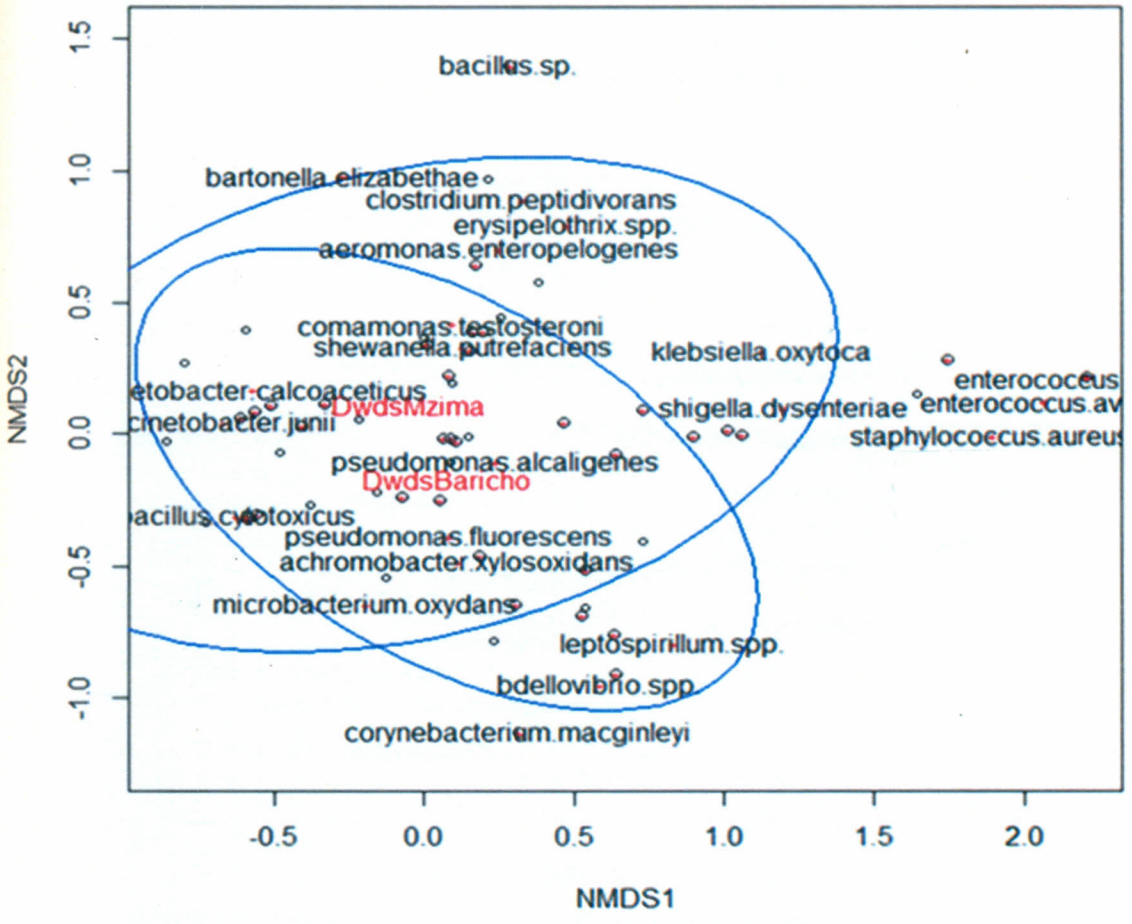


Fig. 4. 28 NMDS of true pathogenic species in Mzima and Baricho water lines

Several pathogenic species were clustered by NMDS with lack of clearly distinct clusters in an excellent representation of the summarized observed distance among biofIm and bulk water pathogens (Stress: 0.063383) in the Baricho water line (Fig. 4. 29). Baricho bulk water had *Mycobacterium, Aeromonas* and *Enterobacter* species. Biofilms on the other hand featured *Agrobacterium, Escherichia, Enterococcus, Legionella, Bdellovibrio* and *Acenotobacter* species. *Pseudomonas* species featured in both water and biofilms.

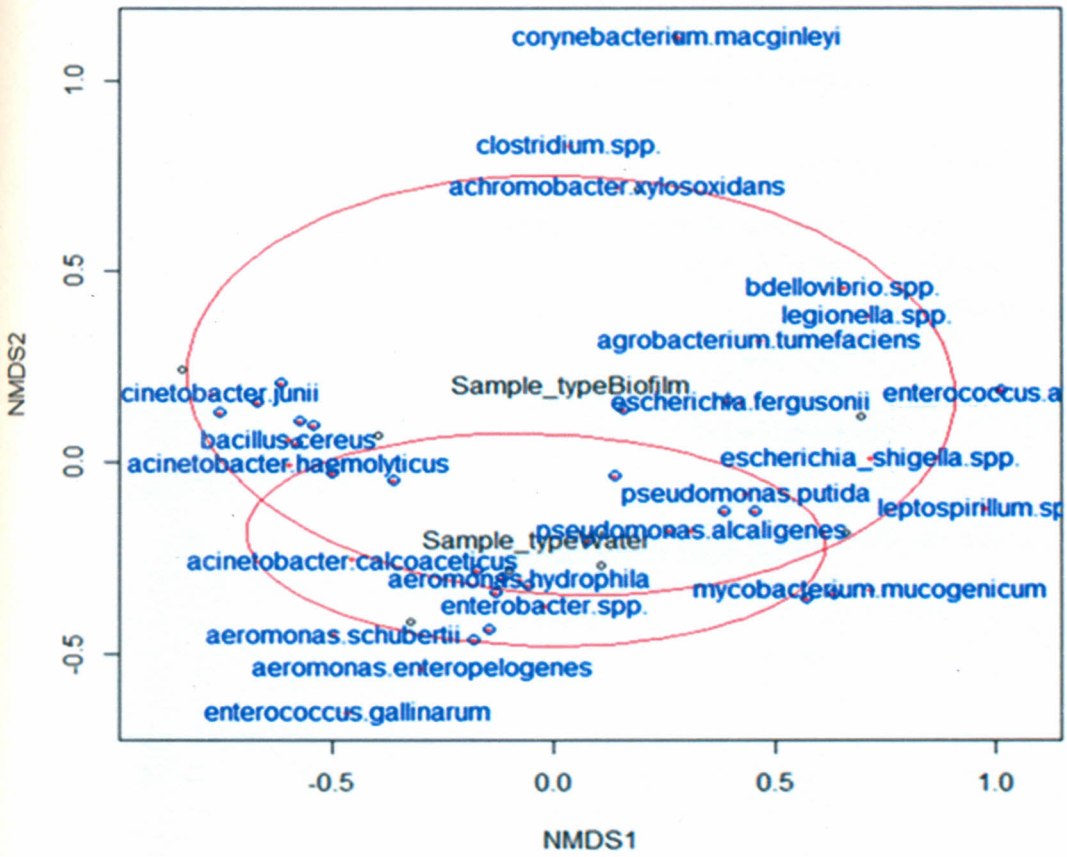


Fig. 4. 29 NMDS of true pathogenic species in the Baricho water line

Ordination of pathogenic species of the Mzima water line yielded two very distinct clusters whose distance is excellently represented (Stress: 0.06591182) with very minimal overlaps. The water samples were uniquely dominated by the *Pseudomonas, Klebsiella, Escherichia, Shigella* and *Bacillus* species (Fig. 4. 30). Mzima biofilms were dominated by *Aeromonas, Bartonella* and *Clostridium* species. However, the *Acinetobacter* species were abundant in both water and biofilms of the Mzima line.

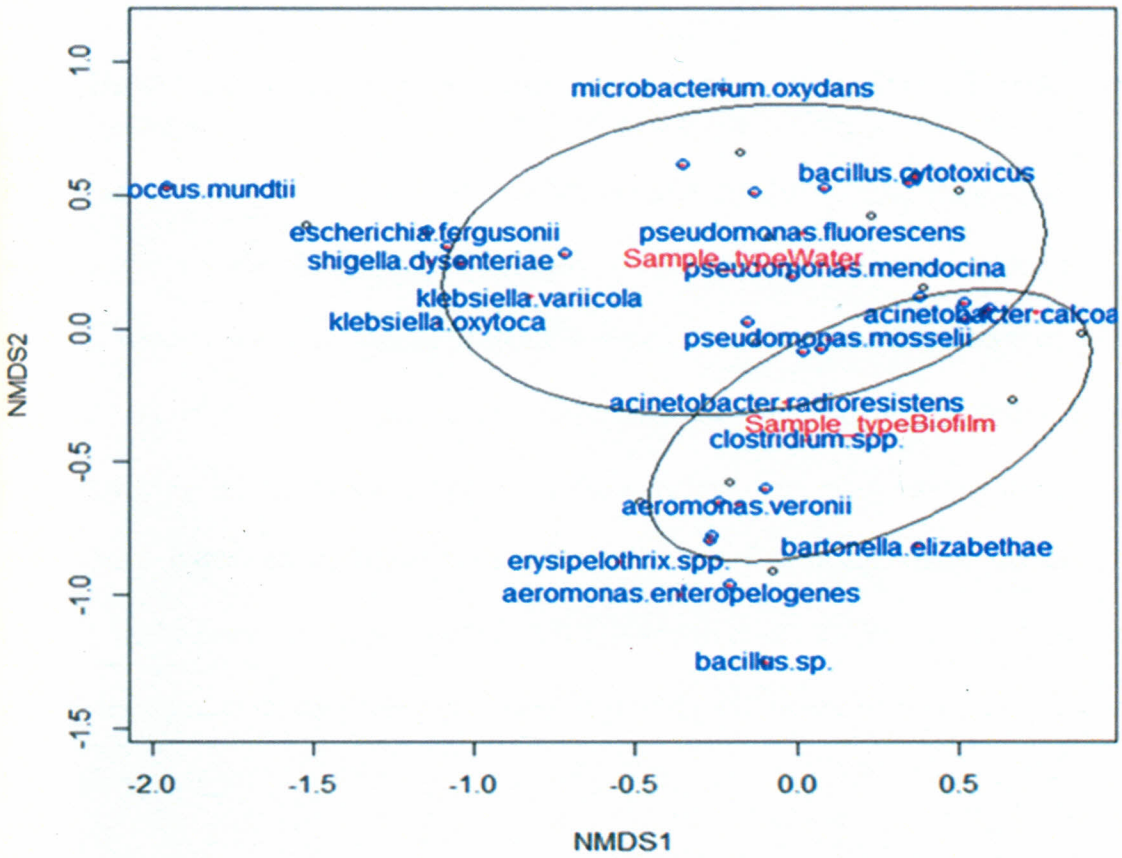


Fig. 4. 30 NMDS of true pathogenic species in the Mzima water line

CHAPTER 5: DISCUSSION

5.1 Bacterial Community Composition and Diversity in the Water Distribution Systems

Pyrosequencing of the 21 water and biofilm samples yielded 27,937 sequences, from which 2,294 unique OTUs were identified. A previous pyrosequencing survey of drinking water meters in Urbana Illinois revealed a similarly high number of total bacterial OTUs (Hong *et al.*, 2010). Similarly, the number of sequences from pyrosequencing is also much greater than those reported in other studies using 16S rRNA gene based conventional molecular analyses of drinking water, which are typically in the hundreds (Revetta *et al.*, 2010; Poitelon *et al.*, 2009). This confirms the great potential of pyrosequencing for assessment of the full extent of microbial diversity in drinking waters.

A total of 20 bacterial phyla and 6 candidate phyla were identified from pooled samples and were dominated by the *Proteobacteria* (73.2%), followed in abundance by *Firmicutes* (13.4%), *Bacteriodetes* (5.9%), *Actinobacteria* at 2.3% while *Planctomycetes* made up1.2% of the total bacterial phylotypes. Candidate divisions contributed 0.9% of the phyla (Fig. 4.3) while only 0.1% were unclassified. Dominance of *Proteobacteria* has been reported for freshwater ecosystems, including drinking water systems (Poitelon, *et al.*, 2009; Kormas *et al.*, 2010; Pinto *et al.*, 2012). Another study in Ohio, using similar techniques revealed the presence of *Proteobacteria, Cyanobacteria, Actinobacteria, Bacteriodetes*, and *Planctomyetes*. However, 57.6% of the sequences could not be classified (Revetta *et al.*, 2010). A fairly exhaustive classification in this study that saw only 0.1 % of the phylotypes not being classified

may be due to the use of the new SILVA database system. This system has been reported to provide a comprehensive, quality controlled, richly annotated and aligned, reference rRNA databases to support the molecular assessment of biodiversity, especially of bacteria (Elmar *et al.,* 2007). The candidate bacterial division *TM7* which was recovered in significantly high proportions in the biofilms from the Baricho water line at source is highly ubiquitous. It has been detected in environments ranging from soils, sediments and wastewater sludge and is also an aetiologic agent for periodontitis and inflammatory bowel disease (IBD) in humans (Dinis *et al*., 2011; Hugenholtz *et al*, 2001).

Previous studies which indicated that *Alpha-*, *Beta-* and *Gammaproteobacteria* were among the predominant bacteria in drinking and mineral water (Falcone-Dias *et al.,* 2012; Manuel *et al.,* 2010; Williams *et al.,* 2004) as was seen in the Baricho line. The study observed dominance of the *Proteobacteria,* similar to previous studies, but with higher propotions of *Nitrospirae* and *Bacillus* as recovered in the Baricho line. Members of the genus *Bacillus* form environmentally resistant spores that can withstand prolonged and constant contact with chlorine allowing their persistence in the water system. The presence of *Nitrospirae* species in the biofilm samples suggests the occurrence of nitrite oxidation in the water lines.

Relative abundance of bacterial taxa was generally high in the two water lines in both the bulk water and biofilms, although the highest abundance was recorded in the Baricho water line. This phenomenon may be due to the poor maintenance of the line

may be due to the use of the new SILVA database system. This system has been reported to provide a comprehensive, quality controlled, richly annotated and aligned, reference rRNA databases to support the molecular assessment of biodiversity, especially of bacteria (Elmar *et al.,* 2007). The candidate bacterial division *TM7* which was recovered in significantly high proportions in the biofilms from the Baricho water line at source is highly ubiquitous. It has been detected in environments ranging from soils, sediments and wastewater sludge and is also an aetiologic agent for periodontitis and inflammatory bowel disease (IBD) in humans (Dinis *et al*., 2011; Hugenholtz *et al*, 2001).

Previous studies which indicated that *Alpha-*, *Beta-* and *Gammaproteobacteria* were among the predominant bacteria in drinking and mineral water (Falcone-Dias *et al.,* 2012; Manuel *et al.,* 2010; Williams *et al.,* 2004) as was seen in the Baricho line. The study observed dominance of the *Proteobacteria,* similar to previous studies, but with higher propotions of *Nitrospirae* and *Bacillus* as recovered in the Baricho line. Members of the genus *Bacillus* form environmentally resistant spores that can withstand prolonged and constant contact with chlorine allowing their persistence in the water system. The presence of *Nitrospirae* species in the biofilm samples suggests the occurrence of nitrite oxidation in the water lines.

Relative abundance of bacterial taxa was generally high in the two water lines in both the bulk water and biofilms, although the highest abundance was recorded in the Baricho water line. This phenomenon may be due to the poor maintenance of the line

leading to frequent breakdown of the break pressure tanks and also vandalization by the public in an attempt to access free water.

With regard to Shannon, Simpson and Inverse Simpson indices, BRW5 recorded the greatest diversity, while BRW6 exhibited the lowest diversity both at phylum and generic levels. Interestingly, both sampling points are household meters at the use point of the water and are therefore accessed more often compared to the main water line. This observation could be contributing to the introduction of bacteria and growth factors in to the system during the day to day anthropogenic activities. Compared to BRW5, BRW6 has detectable levels of residual chlorine of 0.01 mg L^{-1} (Table 4.7) and this may explain the drastic reduction in bacteria proportions.

Sequences corresponding to Methylobacteria were detected in this study indicating the activity of methanogen associated process in the distribution systems. *Methylobacterium adhaesivum,* Candidatus *Methylacidiphilum spp., Methylotenera spp., Methylophilus spp.,* uncultured *Methylotenera* bacteria of freshwater sediment, *Methylomonas spp., Methylosinus spp., Methylibium petroleiphilum, Methyloversatilis spp., Methylosinus sporium* and *Methylocella palustris* were detected in different biofilm samples of the two water lines and BRW5 water. Methylobacteria phylotypes were recovered in two water samplesN (MZW1 and BRW5), and four biofilm samples (BRB 1, 2 and MZB 1, 4) and were particularly abundant in biofilms comparable to bultk water. This may be due to the creation of anoxic cores in the biofilm matrix due to the accumulation and death of bacteriaand archaea. Recent studies of water meter biofilms detected sequences corresponding to the family *Methylococcaceae* (Revetta *et*

al., 2010; Hong *et al.*, 2010). *Methylomonas* sequences have also been detected previously in drinking water (Revetta *et al.*, 2010; Hong *et al.*, 2010). *Methylobacterium*, a genus of methylotrophic bacteria that oxidize methyl compounds such as methanol but cannot metabolize methane, was also one of the most commonly detected taxa in the biofilms, though at low proportions of about 1% of the total sequences. *Methylophilus*, a methylotrophic bacterium, has also been detected in biofilms within drinking water meters (Hong *et al.*, 2010).

The factors favoring high abundance of methanotrophic and methylotrophic taxa within drinking water distribution systems are unclear. High concentrations of methane or methanol in drinking water are not expected, though these compounds could be produced in anoxic sites within water distribution systems via anaerobic processes such as methanogenesis or fermentation. Utilization of haloacetic acid, which is a common by-product of chlorination of drinking water could be another factor favouring abudance. A recent study isolated a *Methylobacterium* strain from a water distribution systems biofilm that was capable of growth with haloacetic acid as the sole carbon source (Zhang *et al.*, 2009).

From the bacterial community information obtained, the potential functions of some species were inferred. Among these species were *Thioclava, Thiobacillus, Thiothrix, Acidothiobacillus, Bosea thioxidans, Thioalkalispira, Blastochoris sulfoviridis* and *Sulfurimonas* which are sulfur-oxidizing bacteria and *Desulfuromonas, Desulfatibacillum alkenivorans, Desulfobulbus, Geobacillus, Geoalkalibacter, Geothrix, Geobacter grbiciae, Geobacter spp, Geobacter bemidjiensis,* and

Desulfovibrio which are sulfate-reducing bacteria. These bacteria may have an important role in sulfur transformation in water. "Sulfur bacteria" or "sulfate-reducing bacteria" can change sulfate and other sulfur containing compounds, including natural organic materials, to hydrogen sulfide gas. Although sulfur bacteria are not harmful, the hydrogen sulfide gas can be hazardous at high levels. Sulfur bacteria also produce slime and can promote the growth of other bacteria, such as iron bacteria. The slime is known to clog plumbing and corrode pipes and other metal components of the water distribution system. Microbial heterotrophic sulfur cycling has been shown to be significant in other environments (Sorokin, 1996). *Flavobacterium* species have been also shown to oxidize thiosulfate (Teske *et al.*, 2000) and dimethyl sulfide (Green *et al.*, 2011) in culture, although organic carbon was still required for growth.

The principal components analysis (Fig. 4.11) indicated that nitrate nitrogen concentration was a significant driver of the composition of the bacterial communities. In the Mzima water line, the source water had high NO_3-N content and an equally high proportion of *Nirospirae*. The ammonia-oxidizing bacteria (AOB) (*Nitrosococcus*) and (NOB) (*Nitrospira*) found in the water and biofilm samples suggested that nitrification possibly occurs in the water lines. The nitrite-oxidizing communities were composed primarily of *Nitrospira,* with higher numbers of the *Nitrospira* occurring in biofilm samples. The presence of *Nitrospirae* species in the biofilm samples suggests the occurrence of nitrite oxidation in the water lines. Previous studies have indicated that the presence of free ammonia is the principal cause of nitrification in DWDS (Wilczak *et al.,* 1996). Nitrification is a two-step biological process that converts ammonia into

nitrite by AOB and further into nitrate by nitrite-oxidizing bacteria (NOB) (Zhang *et al.*, 2009; Regan *et al.*, 2002). Nitrification occurrence is conceptually considered a balance between AOB growth on free ammonia and the inactivation from residual chlorine. Certain nitrobacteria (*Nitrospira, Nitrosococcus, Nitrobacter, Candidatus Nitrotoga spp* and *Pseudomonas nitroreducens*) were found mostly in the biofilm samples. Multiple studies have identified nitrifying bacteria in drinking water distribution systems (Lipponen *et al.*, 2002, Regan *et al.*, 2002). Although nitrobacteria are not considered human pathogens, nitrification can affect drinking water quality and results in regulatory violations.

The *Xanthomonadaceae* which are obligate aerobic chemoorganotrophs were also abundant in the distribution system. This family includes some well-known plant pathogens (Brenner *et al.*, 2005). Organisms from the family were identified both in bulk water and biofilms of the two water lines (*Xanthobacter autrophicus, Pseudoxanthobacter spp,* and *Pseudoxanthomonas mexicana*). Being plant pathogens, the bacteria are likely to be transient in the water system, being introduced in water through plant debris or soil and then getting incorporated into biofilms. These organisms have previously been isolated from drinking water and drinking water pipe biofilms using culture based techniques (Critchley and Fallowfield, 2001). A previous pyrosequencing survey of drinking water biofilms however did not detect family *Xanthomonadaceae* (Hong *et al.*, 2010).

5.2 Relative Abundance of Bacterial Communities in Water Distribution Systems

The study findings revealed that the distributions of bacterial classes among sources and sample types is relatively even (Table 4.4) despite differences in overall richness of the samples. MANOVA analyses revealed that bacterial communities differ with source and sample type, suggesting that the differences in community structure is likely driven at taxonomic levels lower than Class (Family, Order and Genus.). The heterogeneity among samples is demonstrated by the occurrence of only three shared OTUs between the water and biofilms of both the Mzima and Baricho lines. This demonstrates variations in bacterial composition in different sample types that is biofilms and bulk water.

Notably high bacterial genus richness was also recorded in MZB1, BRB1 and BRB2. These three samples actually represent biofilms sampled from the source points of the two water lines. The results confirm that bacterial richness at genus level is greater in biofilm samples and the water lines at sources. BRW5 and BRW6 had the highest and lowest bacterial taxa richness respectively. This could probably be as a result of lack of selective pressure exerted by chlorine since the water is not chlorinated at the point of abstraction. Another possible reason for a greater richness at the source is the particularly high nutrient levels, especially in Mzima water line (Table 4.8) where nitrate and phosphate levels were exceptionally high.

The Renyi profiles (Fig. 4.7) confirm that all samples were low in species evenness. The highest recorded evenness was in MZW5 and BRW6 (1.000). Water therefore has less variation in bacterial community species unlike biofilms samples BRB 1 and BRB2

from which the lowest evenness was recorded (0.19), indicating biofilms are species poor communities. This phenomenon may be as result of the freely available nutrient for bacterial growth in water comparable to the fairly constrained environment in biofilms that would require special adaptations evident only in some bacteria species, for bacterial survival.

Generally, drinking water distribution systems are considered oligotrophic environments (with low contents of nutrients like carbon, nitrogen and phosphorous). However, an increase in nutrient levels in water promotes biofilm formation. Studies have shown a positive relationship between the concentration of nutrients in drinking water and bacterial regrowth in drinking water distribution systems (Owen *et al.,* 1995; Van der Kooij, 1992). Servais *et al.,* (2004) and Batté *et al.,* (2003) observed that the addition of phosphorus did not affect the accumulation of biofilm although phosphorus was being incorporated into the biofilm. Other researchers observed that the limiting nutrient was phosphorus in drinking water distribution systems in Japan (Sathasivan and Ohgaki, 1999) and Finland (Lehtola *et al.,* 2002). The detection of the limiting nutrient in drinking water distribution systems is very important since the addition of phosphate based compounds has been proposed to prevent pipe corrosion and the bacterial regrowth (Appenzeller *et al.,* 2002). The concentration of nutrients such as carbon/nitrogen ratio is important to the production of extracellular polymers that affects the adhesion of microorganisms to surfaces (Veiga *et al.,* 1997). Hydrodynamic conditions and nutrients are the two main parameters that influence biofilm growth in particular the biofilm structure, density and thickness (Horn *et al.,* 2002; Wasche *et al.,*

2002; Wimpenny and Colasanti, 1997). High shear stress and nutrient limitations lead to thin and dense biofilms that will have reduced internal nutrient diffusion (Melo and Vieira, 1999; Zhang and Bishop, 1994) and increased resistance to removal and cohesion (;Paris *et al.,* 2007; Chen *et al.,* 2005; Van Loosdrecht *et al.,* 1997; Vieira *et al.,* 1993).

In terms of abundance, MZW7 had the highest bacterial abundance (897 phylotypes) while MZW5 had the lowest (200 phylotypes) (Table 4.4). MZW7 was sampled from a water meter at a user point, a community water kiosk, while MZW5 was sampled from a household meter in close proximity to the chlorination tank, where illegal and cross connections with the potential of introducing contaminating bacteria are absent.

Nonmetric dimensional scaling (Fig. 4.8) revealed the clustering of bacterial phyla by water lines to show how the observed phylogenetic distances of the phyla is distributed in the two water lines. The ordination ellipses show distinct clusters of phyla with obvious overlaps in a large number of phyla between the two water lines.This implies that though most phyla are unique to one water line, some can be found in both water lines in similar proportions. There was dissimilarity in phyla composition of the two water lines (Mantel correlation test $r = 0.27$, $p = 0.001$) despite being operational for more than 50 years in almost similar climatic conditions. Both lines convey abstracted ground water whether springing naturally or pumped from boreholes. These lines have therefore attained a level of equilibrium in the microbial processes and interactions that lead to generation of the bacterial biomass in the systems by virtue of being operational for more than 50 years.

Redundancy analysis also confirmed that there was a significant difference in bacteria taxa diversity of Mzima and Baricho water lines (r^2=0.1902, F=4.228, p=0.01). Similarity percentages (Table 4.5) confirmed this dissimilarity and the phyla that contributed significantly to dissimilarity in the two lines were Bacteria *TM7, TM6* and *NPL-UPA2*, and *OD1, OP3*, *BRC1, WS3* and *OP11* Candidate divisions. ANOSIM analysis confirmed that the bacterial diversity in the two water lines was not similar (p>0.01). The depth of genome characterization may be the key contributor to detection of this dissimilarity.Use of pyrosequencing allowed for the identification of these candidate phyla, which would otherwise not be identified by culture dependent techniques and which notably contributed to the dissimilarity.

Predominance of the *Gammaproteobacteria* population over the *Alpha-* and *Betaproteobacteria* populations in most of the samples was evident in this study finding. In contrast, previous studies have reported the predominance of *Alphaproteobacteria* or *Betaproteobacteria* or both in the drinking water (Williams *et al.*, 2004). This could be attributed to the differences in the source water quality (borehole water from Baricho line, versus the spring water of Mzima line), physico-chemical variables of the water, and efficiency of disinfection treatment (Poitelon *et al.*, 2010).

5.3 Physico-chemical Characteristics of Water in Water Distribution Systems

The environmental variables of the water such as pH, temperature and nutrient levels (Table 4.7) from which bacterial sequence reads were obtained provide useful

information on water quality or the potential of water to support bacterial growth. All Baricho water samples had high levels of iron above the acceptable limits. The source of the high amounts is likely to be the metallic pipes used for water transport. Noticeable during sampling was the extensive rusting of the metal rungs in the chlorination and break pressure tanks of the Baricho water line. Manganese and lead levels were also above acceptable limits in the Baricho water line, while Mzima water samples 3, 4 and 7 also had lead levels above the acceptable limits.

Another aspect that may be contributing to high levels of metals in water is the leaching of volatile components from pipe materials that can be metabolized by biofilm microorganisms. Van der Kooij *et al.* (1995) also observed that the polymeric materials in contact with drinking water could release biodegradable compounds, thus enhancing biofilm formation. Corroded iron pipes may offer numerous bacterial attachment sites and bacteria protection from the effect of flow rate and of disinfectants as well as may release undesirable products to the water (LeChevallier *et al.,* 1993). Also, the corrosion products may retain nutrients (such as, humic matter) for subsequent utilization by biofilm bacteria (Butterfield *et al.,* 2002). The corrosion on metallic surfaces may be induced by the activity of physiologically diverse microbial species within the biofilms (Beech, 2004).

Whereas both Mzima and Baricho water lines supply ground water, the Mzima water is abstracted from six spring eyes while the Baricho water is abstracted from seven boreholes sunk on the bank of Sabaki River. The Mzima intake is in a game reserve which is a forest area characterized by trees and thickets which serve as habitats for

wild animals. This may account for the low water temperatures at source due to the shade effect and also for the high levels of nitrates and phosphates as a result of introduction of organic matter into the water from the surrounding environment. Although the Baricho water is from an underground source, it is important to note that refill may occur from the Sabaki River, which may be the source of high levels of lead and manganese. High lead levels may also have been as a result of pump replacement and on going repair work of the water line.

The single occurrence of zinc in BRB2 may have resulted from the introduction of zinc during repairs of the water line since the sample was collected two months after the laying of a new pipe joint and so far no other sample recorded detectable levels of zinc in water. Probably the pipe linings or some of the sealants used may have incorporated in them some zinc compounds. A significant difference in the pH, nitrates, iron and manganese levels between the two water lines is probably as a result of the bioleached metals and nutrients accumulated in the water lines as well as the bacteria that have ability to utilize these nutrients.

Principle Components Analysis (Fig. 4.11) showed that pH, phosphates and nitrates were the critical determinants of variations in bacterial composition in the water lines. Principal Coordinates Analysis (Fig. 4.12) revealed that the most similar samples were those of Baricho biofilms followed in order by Mzima water samples and Mzima biofilms. Samples with the greatest distance or dissimilarity were Baricho water samples. One unique factor about Baricho water samples is their residual chlorine values which are varied along the different phases of the system. Chlorination therefore

may be the single most critical determinant of the variation in the composition and abudance of bacteria in Baricho water samples. Biofilm samples are fairly similar due to stabilization over time as the films mature giving rise to a more or less similar matrix and flora composition. Also the nature of the biofilm hinders penetration of disinfecting agents contributing further to the compositional stability.

Canonical Correspondence Analysis (Fig. 4.13) showed that the physico-chemical variables that significantly affected bacterial diversity and abundance in the two water lines were temperature, iron and nitrates. Iron is probably introduced in the water from corroded and rusted water pipes and metal rungs in the water storage tanks, while nitrates are from organic matter which in the case of the Mzima water line, was abundant at the spring eyes which are located in Tsavo forest. The microbial composition of drinking water distribution systems communities is influenced by several factors and reflects the microflora characteristics of the raw water source (Eichler *et al.,* 2006). Previous studies have shown that distribution system pipe material, temperature, the level of organic carbon available, velocity of water and the disinfectant used in a system are among the factors that may impact the growth and community structure of drinking water distribution systems biofilms (Williams *et al.,* 2004; Camper *et al.,* 2003; Percival *et al.,* 1999).

Median pH values of water in distribution systems regardless of the type of source were close to neutral (7.84). The highest pH value recorded was in Baricho water at source, while the lowest pH values (7.65) were recorded among Mzima water samples (MZW1). Baricho water line recorded pH values as high as 8.12 with a median of 7.95

(Table 4.8) with the genera Paracoccus, Alcaligenes, Staphlococus and Limnobacter being recovered in samples with highest pH values. Higher pH values in Baricho water are probably due to groundwater exposure to salts and metal ions leading to a rise in pH. The proposed WHO guideline for pH is 6.5 - 9.0 while the Kenya Bureau of standards (2005) recommends a pH of 6.5 – 8.5 for drinking water. As such, all of the water samples meet the KEBS standards for drinking water. Other studies have shown that pH strongly influences bacterial community structure at large geographic scales or artificial gradients (Rousk *et al.,* 2010). Extremes of pH (less than 6.5 or greater than 9) can be toxic to aquatic organisms. Eye irritation and exacerbation of skin disorders have been associated with pH values greater than 11. Although pH has no direct impact on consumers, it is one of the most important operational water quality parameters (WHO, 2003).

Absence of a significant relationship between the abundance of bacteria within the biofilms and the concentrations of residual chlorine suggests that low concentrations of chlorine disinfectants may not be effective in limiting biofilm growth within water distribution systems. This can be attributed to the protection provided by the biofilm matrix, as has been demonstrated by previous studies (Liu *et al*., 2011; LeChevallier *et al*., 1988; Ridgway and Olson, 1982). Studies have also demonstrated the influence of disinfectants on the *Proteobacteria* populations where *Alphaproteobacteria* predominated in both chloraminated and chlorinated water (Williams *et al.,* 2004), while *Betaproteobacteria* and *Gammaproteobacteria* in drinking water biofilms were favored by increased chlorination (Mathieu *et al.,* 2009). According to Williams *et al.,*

(2004), following exposure to free chlorine and monochloramine, *proteobacteria* was the predominant phylogenetic group observed in the treated distribution water, suggesting that these organisms are well suited to survive in potable water supplies. Conversely, *Proteobacteria* were found to be more abundant in chloraminated water than in chlorinated water. In another study, Emtiazi *et al.,* (2004) revealed that *Proteobacteria* were also abundant in biofilms of non-chlorinated DW. These studies indicate that microbial community diversity is impacted by the disinfection strategy. There is also evidence that greater bacterial diversity can affect disinfection efficacy and pathogen survival. Simões *et al.* (2010) and Berry *et al.* (2006), provide experimental evidences on the role of the microbial diversity of drinking water isolated bacteria biofilms in their resistance to chlorine disinfection.

5.4 Biofilm and Bulk Water Bacterial Communities

Comparison of bulk water and biofilm microbiomes yielded very interesting findings. Despite the large numbers of OTUs observed in bulk water and pipe biofilms, these communities were not very diverse, with all samples having Shannon index scores of less than two. These low diversity values is a reflection of dominance of these communities by a small number of taxa. Generally biofilms had higher bacterial phyla and generic richness and abundance (Fig. 4.14). This can be attributed to the ability of biofilms to allow microbes to persist and grow in hostile environments such as water distribution systems under low nutrient conditions. In this protected environments, extrapolymeric substances (EPS) provide a diffusional barrier against disinfectants and other deleterious chemicals. Hence, EPS increases the resistance of biofilms and any

detached biofilm clusters to chlorine (Xue *et al.,* 2013b). Using model water distribution system, it was found that during the initial stages of biofilms formation the 16sRNA sequences in biofilms are similar to those found in bulk water. However, DNA and RNA based fingerprints of 20 year old biofilms have shown higher diversity in biofilms communities than in bulk water (Henne *et al.,* 2012). This could be due to the protective barrier generated in biofilms by extrapolymeric substances, nutrients and microbial cells. However, its still unclear the effects of surface materials on biofilm development when polymeric materials such as plastics and metallic materials are compared. Some previous reports have demonstrated that drinking water biofilms grew less on polymeric materials than on iron matrices (Chang *et al.,* 2003; Hem and Skjevrak, 2002; Niquette *et al.,* 2000; Kerr *et al.,* 1999). This observation was attributed to iron corrosion products that favor biofilm protection from mechanical and chemical stresses. Other studies have reported higher biofilm formation on PVC and PE surfaces than on galvanized steel materials (Bachmann and Edyvean, 2006; Cloete *et al.,* 2003; Schwartz *et al.,* 1998). It is important to note that most of the distribution system flow pipes are made of PVC, especially those leading from storage and chlorination tanks to the households or other userpoints.

Alphaproteobacteria dominated in the bulk water while *Beta* and *Gammaproteobacteria* (which include many primary and opportunistic pathogens) were predominant in biofilms. Examination of biofilms from household water meters showed that the major bacteria genera were *Sphingomonas (Alphaprotebacteria) Acidovorax, Methylophilus, (Betaproteobacteria)* and *Lysobacter (Gammaproteobacteria*).

Examination of the Ann Harbor drinking water by 16S rRNA gene sequencing also showed that most of the bacteria were *Proteobacteria* and belonged to the genera A*cidovorax, Variovorax, Sphingopyxix, Ralstonia* and *Novosphingobium* (Lee *et al.*, 2010).

Bulk water on the other hand featured greater species evenness (Table 4.9). Both biofilms and bulk water samples are dominated by the *Proteobacteria*, especially of the *Beta* and *Gamma* type. However, Baricho water had very high levels of *Nitrospirae* (20.2%) (Fig. 4.16) which was were very low in biofilms (0.8%) (Fig. 4.17). The reverse is true for *Alphaproteobacteria* and *Flavobacteria* which are abundant in biofilms (29.6%, 0.4%) but low in water (6.4%, 2.6%) respectively. Bacteria generally found in water distribution systems include *Pseudomonas, Acinetobacter, Flavobacterium, Alkaligenes, Aeromonas, Moraxacella, Enterobacter, Citrobacter, Sphingomonas, Klebsiella, Burkholderia, Xanthomonas, Methylobacteria* and *Bacillus*. These findings are in line with previous studies, which also identified these genera in drinking water distribution systems (Berry *et al.*, 2006; Block *et al.*, 1997). The taxonomic analyses of the bulk water samples in this study generated findings similar to those of the analyses of two former studies of the city of Braunschweig, Germany drinking water distribution systems (Kahlisch *et al.*, 2012; Eichler *et al.*, 2006). Notable are the 3% of all classes which are uniquely uncultured directly and the *Deltaproteobacteria* were recovered only in the biofilms and not from the bulk water samples. Drinking water biofilms are well known to carry diverse microbial communities. Many of the microbes identified in this work are indeed very closely

related to typical drinking water or fresh water microbes, and their presence in biofilms has been described before (Poindexter *et al.*, 2000; Ribas *et al.*, 2000).

Redundacy analysis indicated that *Firmicutes, Actinobacteria, Acidobacteria, bacteria NPL- UPA2* and *Candidate division OD1* contributed greatly to the dissimilarity between biofilms and bulk water microbiomes. This scenario was also evident in studies of diversity of biofilms and bulk water in the individual water line. The phyla causing dissimilarity are generally culturable with only one candidate phylum, unlike the comparison of the two water lines where majority of the phylotypes contributing to dissimilarity belonged to candidate phyla.

The study findings suggest that in mature water distribution system biofilms, as is the case with both the Mzima and Baricho biofilm sample, bacterial communities generally showed similarity in structure. These observations are corroborated by Martiny *et al.*, (2003), who showed in their model water distribution system that after 3 years most biofilms from different sampling positions clustered together, to form a homogeneous community structure. With respect to the diversity and structure of biofilm communities, there are similar mechanisms employed by microbes in structuring these communities in water distribution systems. This behavior suggests that all biofilms provide a similar number of niches but are filled with different species. During the initial stages of biofilm development microcolonies are usually formed following the attachment of single cells to the pipe surface. Species diversity increase thereafter and the biofilms formed may reach steady state and high ecological diversity, comprising both heterotrophs and autotrophs, only after 2-3 years (Martiny *et al.*, 2003; Berry *et al.*,

2000). A rapid decrease of abundant species with a long "tail" of rare species in drinking water biofilms was also observed in a recent study using pyrosequencing (Hong *et al.*, 2010). The current model of abundant and rare members describes the bacterioplankton community in pelagic ecosystems, consisting of a core community with few taxa that are highly abundant and a seed bank with nearly infinite numbers of low-abundance phylotypes (Höfle *et al.*, 2008; Huber, *et al.*, 2007; Pedrós-Alió, 2006). The Mzima and Baricho water distribution lines have been operational for over 50 years guaranteeing that the biofilms are truly representative of mature water distribution systems biofilms.

Previous studies have indicated that some bacteria, which might be resistant to chlorination at a certain degree, could survive after disinfection and proliferate in the water distribution systems (Wang *et al.*, 2010; Norton *et al.*, 2004). This is consistent with the findings of the present study, which strongly suggested that the bacterial community in BRW 6 a sampling point that occurred after chlorination had an increase in diversity comparable to the prior sampling points which had detectable residual chlorine in the water line. The high diversity may be due to poor efficiency of residual disinfectants in drinking water to inactivate pathogens (Payment, 1999); especially considering the oligotrophic conditions in the drinking water distribution system networks. Studies by Poitelon *et al.*, (2010) showed that the bacterial community composition changed following chlorine disinfection as *Alpha, Beta* and *Gamma-proteobacteria* decreased after chlorination. It is also important to note that

Baricho water from sampling point 6 is located at community level and anthropogenic activities may also be contributing to the higher bacterial diversity and abundance.

In biofilms of both water lines, a few methanotrophs belonging to the family *Methylococcaceae* of the *Gammaproteobacteria* were detected. Methanotrophs can play a role in the nitrogen cycle through mechanisms such as their coexistence with heterotrophic nitrifiers via cross-feeding of metabolites (Modin *et al.*, 2007). In the presence of environmental gradients of oxygen and methane, methanotrophs can also support denitrification by both creating anoxic environments conducive for denitrification during methane oxidation and releasing organic substances that can be used by coexisting heterotrophic denitrifiers (Modin *et al.*, 2007). The association between the two groups of microorganisms has led to the suggestion of using methane as an external carbon source for biological denitrification of nitrate contaminated waters (Modin *et al.*, 2007).

Unlike Mzima water line, Baricho water line had a relatively high abundance of members of the genus *Methylophilus,* especially in its biofilms. The genus *Methylophilus* consists of a group of bacteria that oxidize methyl compounds but not methane. An abundance of *Methylophilus* bacteria in cave water that receives a high flux of methane has been reported (Chen *et al.,* 2009b). It has been suggested that the methylotrophs possibly fed on methanol produced by the methanotrophs during methane oxidation (Chen *et al.,* 2009b; Qiu *et al.,* 2009) or that the methylotroph-like sequences were in fact from as yet uncultured methanotrophs in the *Betaproteobacteria* (Chen *et al.,* 2009b). Since members of the genus *Methylophilus* can use nitrate as a

nitrogen source (Jenkins *et al.,* 1987), it is also possible that the *Methylophilales* detected in Mzima and Baricho water and biofilms can carry out denitrification using the methanol supplied by the methanotrophs. Methanol dependent denitrification by members of the family *Methylophilaceae* has recently been reported suggesting a potential role of this group of bacteria in both nitrogen and carbon cycling (Kalyuhznaya *et al.,* 2009).

The genera *Sphingomonas* and *Acidovorax* recorded in both the Mzima and Baricho lines were more predominant in biofilms as compared to water. These two genera comprise of metabolically diverse species capable of using a wide range of naturally occurring compounds. With respect to the water distribution systems, members within these two genera have been implicated in copper pipe corrosion (Critchley *et al.,* 2004) and demonstrated to enhance biofilm formation by other bacterial groups (Simoes *et al.*, 2007). Members of these two genera could be potential opportunistic pathogens (Koskinenn *et al.*, 2000; Zanetti *et al.*, 2000). In contrast to bulk water, the majority of biofilm phylotypes were considered to be of soil, sludge or sediment origin. Members of the key genera *Rhizobiales*, *Nitrospira*, and *Thiobacillus*, which were found in drinking water biofilms, are known to contribute to the biogeochemical cycling of nitrogen or sulfur and their presence is indicative of their active participation in the ammoniam decomposition component of the nitrogen cycle.

5.5 Patterns of Variation of Bacterial Communities in Different Phases of the Water Distribution Systems

As the pipeline runs its course, significant changes could be seen in phyla composition both in diversity and proportions of operational taxonomic units (Table 4.1). *Proteobacteria* dominated Baricho water at endpoint (100%). A total of 8 phyla were present at the source only but not at endpoints, most of which occurred in biofilms only. These were *Nitrospirae, Elusimicrobia, Cyanobacteria, Gemmatimonadetes, NP-UPA2,* and Candidate Divisions *TM7, OP11* and *OD1. Cyanobacteria* have previously been found in drinking water (Revetta *et al.*, 2010). It is suspected that they may have found their way into the distribution system from the open break pressure tanks. It is possible that some of the cyanobacterial populations detected in this study can temporarily survive in the darkness under anaerobic, reducing conditions. It is also possible that they are only trapped in the biofilms and are not metabolically active. Some species of cyanobacteria are considered contaminants in drinking water, since they can produce toxins.

Mzima water at source was dominated by *Betaproteobacteria* (28.65%) followed by *Alphaprotobacteria (*24.3%) and *Gammaproteobacteria* at 15%. The endpoint water registered a a high increase in *Betaproteobacteria* (67%) followed by *Gammaproteobacteria*, which increased to 21%. Notably absent were *Alphaproteobacteria. Clostridia,* which contributed 8.4% of the class phylotypes was present in end point water but not at the source. As the water gets subjected to chlorination and the shear pressure increases as water flows gravitationally or through

pumping, the biofilms may slough off and some bacteria may die off from the effect of the chlorine.

Mzima biofilms were predominated by *Alpha, Beta and Gammaproteobacteria* both at source and endpoint although the proportion of the *Gammaproteobacteria* increased significantly between the source (3.6%) and endpoint (22.5%). Previous studies have reported the predominance of *Alphaproteobacteria* or *Betaproteobacteria* or both in drinking water distribution systems (Williams *et al.*, 2004). In contrast, the findings of this study revealed the predominance of the *Gammaproteobacteria* population over the *Alpha-* and *Betaproteobacteria* populations in most of the samples. This difference can be as a result of differences the molecular methods applied, variations in water quality, the nature of pipe materials (Poitelon, *et al.*, 2010) and the growth stages and the ages of the studied biofilms. Pyrosequencing and blasting against SILVA and Greengenes databases have been shown to yield variations in taxa recoveries (Elmar *et al.*, 2007). Studies have also demonstrated a greater adverse influence of disinfectants on the *Proteobacteria* populations leading to the domination by *Alphaproteobacteria* in both chloraminated and chlorinated water (Williams *et al.*, 2004). Similarly, the presence of *Betaproteobacteria* and *Gammaproteobacteria* in drinking water biofilms has been shown to be favored by increased chlorination (Mathieu *et al.*, 2009).

Another notable observation was the increase in class *Clostridia* from 2.2% in biofilms at source to 20.9% in endpoint biofilms. Interference by chlorine at the source is lacking due to the lack of chlorination and the high flow velocity may also contribute to this increase. Biofilm density and detachment increase with the flow velocity (Melo, 2003).

Higher flow velocities increase the cellular hydrophobicity promotes cell aggregation and hence biofilm accumulation. Under low flow velocities and in the presence of nutrients, biofilms grow quickly with a low dense structure but with many pores, channels and protuberances (Van Loosdrecht *et al.,* (1995). Nutrient transport rates within the biofilm increase with the flow velocity until a maximum value is reached, and then decrease as the velocity is further increased. This transport rate promotes bacterial growth within the biofilm. Research on effects of flow velocity on biofilm accumulation has yielded some controversial results. A larger majority of studies have concluded that biofilm formation increases with flow velocity (Paris *et al.,* 2007; Simões, 2006; Lehtola *et al.,* 2006: Percival *et al.,* 1999), others have come with contrary findings (Chen *et al.,* 2005; Tsai, 2005 Soini, 2002). Several studies focusing on biofilm growth have provided a mechanistic explanation of the effects of hydrodynamics on biofilm growth (Liu *et al.,* 2003; Liu and Tay, 2001). This may also account for why source biofilms having a greater bacterial diversity with more than 20 classes represented while endpoint biofilms featured around 10 different bacterial classses.

Intakes at the source encounter minimal disturbance from man and are likely to have developed a stable microflora comparative to the other phases of the distribution systems. This is evident in the Baricho water which had a noticeably low number of bacterial classes. Endpoint water had a characteristically low diversity represented by two classes; *Betaproteobacteria* (46.2%) and *Gammaproteobacteria* 53.8%. Source water was also predominated by *Gammaproteobacteria* (60%) while *Beta* and

Alphaproteobacteria were 9.1% and 12.7% respectively. *Bacilli* (12.7%) were present only in source water but not water at the endpoint. Baricho biofilms also displayed a greater bacteria class diversity at source, comparable to the end point although most of the classes had very low abundance values of less than 1%.

Higher diversities at source may also be as a result accumulated sediemnts in the water collection tanks. Such sediments can consist of either organic matter, including micro-organisms, or insoluble material, mainly iron and manganese. Significant microbial activity may occur in accumulated sediments. Organic and inorganic particles can also accumulate in low-flow areas or dead-ends of the drinking water distribution systems, and enhance microbial activity by providing protection and nutrients (EPA, 2002). Biofilms that slough can accumulate in the periphery of distribution systems leading to sediment accumulation and the proliferation of some microorganisms (Van der Kooij, 2000). Sediment accumulation may also lead to decrease of disinfectant residual in water resulting in lower bacterial diversities along the distribution lines.

Even so, regrowth may occur when the residual chlorine decays further down in the distribution system (Chandy and Angles, 2001). This may account for the high bacterial composition and diversity at distribution systems endpoints. Some studies have demonstrated that chlorine is able to control biofilm formation by reducing the rate of biofilm growth, promoting the biofilm detachment and decreasing the activity of microorganisms (; Codony *et al.,* 2005; Codony *et al.,* 2002; Lund and Ormerod, 1995; De Beer *et al.,* 1994). However, the presence of residual chlorine is also one of the stress factors that lead to biofilm formation (Kokare *et al.,* 2009).

Interestingly, *Agrobacterium tumefaciens* an organism extensively exploited in genetic engineering for its potential to initiate faster and gigantic growth of cells was recovered in both water line sources, raising queries on the implications of lateral gene transfer within the water distribution systems. Studies have shown that *Agrobacterium tumefaciens* is one of the viable but nonculturable bacterial communities that can survive in water (Byrd *et al.,* 1990). The implication is that there is a possibility that being a soil organism it may get into drinking water at the soil water interfaces but probably was not detected using culture dependent techniques.

The taxonomic analyses revealed that drinking water treatment significantly influence the microbial structure, as indicated by the large drop of several dominant bacterial phyla in BRW 3 after treatment except *Proteobacteria* (Fig. 4.20). The dominant classes in *Proteobacteria* obviously shifted from *Alpha* and *Betaproteobacteria* in Mzima water and biofilms at source respectively (Fig. 4.21) to *Betaproteobacteria* and *Gammaproteobacteria* at Mzima endpoint water and biofilms respectively. A general shift in dominance from *Gammaproteobacteria* at source to *Alphaproteobacteria* and *Betaproteobacteria* at endpoint in the Baricho water line suggests that *Alphaprotebacteria* and *Betaproteobacteria* are tolerant to chlorination which occurs during water treatment. A study by Koskinen *et al.* (2000) reported that *Alphaproteobacteria* can persist after chlorination. The survival of *Sphingomonadaceae,* a family under *Alphaproteobacteria* has been associated with its high resistance to chlorination. Thus, the bacteria in *Sphingomonadaceae* family are

often abundantly found in drinking water systems (Vaz-Moreira *et al.*, 2011; Srinivasan *et al.*, 2008).

Acinetobacter which was present in all sampling points is a genus of Gram-negative, heterotrophic bacteria (Brenner *et al.*, 2005) that is commonly found in soils and groundwater (McKeon *et al.*, 1995; Shirey and Bissonnette, 1991). *Acinetobacter* is also one of the most common groups of bacteria usually isolated from drinking water (Szewzyk *et al.*, 2000). A number of *Acinetobacter* species have been shown to produce biofilms (Hansen, 2007; Tomaras *et al.*, 2003). Hence the presence of *Acinetobacter* in the pipe biofilms was not surprising. However, the dramatic variation in *Acinetobacter* abundance between source and endpoints samples was remarkable with higher proportions being recorded at the end points. Although bacteria from the genus *Acinetobacter* are generally aerobes, there are some species within the genus that can utilize nitrate as an electron acceptor when oxygen is not present (Wentzel *et al.*, 1986), allowing them to thrive in anoxic environments like water conveying pipelines, there by accounting for the *Acinetobacter* increase at endpoints.

5.6 Pathogenic Species in Drinking Water Distribution Systems

According to the guidelines for drinking water quality, there is no tolerable lower limit for pathogens in water intended for consumption, preparing food, drink or for personal hygiene. Hence drinking water should not contain human pathogenic (KEBS, 2005; WHO, 2003). In the present study, at least 140 species of potentially pathogenic bacteria were identified in the water distribution system. Opportunistic pathogens

present a unique challenge in drinking water systems. Waterborne opportunistic pathogens, including *Legionella pneumophila, Mycobacteria* (NTM), *Pseudomonas aeruginosa,* and *Acanthamoeba spp.*, have become an emerging public health concern (September *et al.*, 2004). Several genera known to carry pathogenic strains were detected in the water samples by the pyrosequencing analysis (Fig. 4.26). *Escherichia/Shigella* was recorded in the source water in both water lines at different proportions. Some *Staphylococcus, Pseudomonas,* and *Enterococcus* species were also detected in the water samples. The high levels of pathogenic species in drinking water may indicate that either disinfection is not effective or the bacteria are developing mechanisms of resistance to the disinfectants. In this study particularly, there was notably some occasion when chlorination was not carried out due to administrative challenges. Past studies have cited *Legionella* and *Mycobacterium* as the most commonly detected pathogenic bacteria in drinking water systems and that the two are typically present in trace quantities (Holinger *et al.*, 2014; Revetta *et al.*, 2013; Marciano-Cabral *et al.*, 2010; Feazel *et al.*, 2009;Vaerewijck *et al.*, 2005).

Generally the number of truly pathogenic species in Mzima water line was not different at source and endpoint, but there was a very significant reduction in pathogenic species in Baricho water line endpoints comparable to the source. Baricho source water and biofilms featured 31 and 24 species respectively, while endpoint water featured 20 species and biofilms 15 species. This could be as a result of the effect of chlorination at BRW4 where there is a reduction in pathogenic species. More *Mycobacterium,*

Pseudomonas, Acinetobacter and *Aeromonas* species were found in the Mzima and Baricho water lines compared to previous studies.

Past studies have reported that *Acinetobacter*, *Aeromonas*, *Alcaligenes*, *Bacillus*, *Burkholderia*, *Citrobacter*, *Enterobacter*, *Flavobacterium*, *Klebsiella*, *Methylobacterium*, *Moraxella*, *Pseudomonas*, *Serratia*, *Mycobacterium, Sphingomonas and Xanthomonas,* which are Gram negative, and *Staphylococcus* and *Arthrobacter/Corynebacterium* which are gram positive, are the predominant bacterial genera in drinking water distribution systems (Berry *et al.,* 2006; Block *et al.,* 1997). The study findings indicated that Gram-negative bacteria are usually predominant over Gram-positive bacteria while *Pseudomonas* species are the most abundant bacterial organism in supply systems, regardless of the water source.

Some genera such as *Acinetobacter, Stenotrophomonas*, *Burkholderia, Flavobacterium, Pseudomonas, Staphylococcus* and *Escherichia/Shigella* are known for their facultative pathogenic and nosocomial character (Das *et al.,* 2002). Gram-negative pathogens such as *E. coli, Klebsiella, Pseudomonas, Serratia, Enterobacter, Edwardsiella* and *Bacillus* species were also identified. Although they can cause facultative infections in those with certain predispositions, they can also live in symbiosis with plants or can be used as pro- and prebiotics for both plants and humans (Ryan *et al.,* 2009).

Metagenome analyses confirmed the ubiquity of mycobacteria in drinking water distribution systems. Other microbial diversity studies have documented the presence of mycobacteria in drinking water distribution systems (Covert *et al.*, 1999). The genus

Mycobacterium consists of approximately 100 species, including a large number of species that are either non-pathogenic or pathogenic under certain situations (Vaerewijck *et al.,* 2005). Mycobacteria have been detected previously in drinking water distribution systems via culturing with *M. gordonae* and *M. intracellulare* being the most frequently detected *Mycobacterium* species (Pryor *et al*., 2004). *M. gordonae*, which is among the most frequently reported mycobacteria in drinking water and water distribution systems (Vaerewijck *et al.,* 2005), is generally considered non-pathogenic (Weinberger *et al*., 1992). In this study, the Mycobacteria species of pathogenic value were identified as; *M. mucogenicum, M. brumae, M. senegalense, M. duvalii, M. mageritense, M. asiaticum, M. gadium, M. laticola* and *M. poriferae.* Several recent studies have detected species related to *Mycobacterium* in chlorinated drinking water (Gomez-Alvarez *et al*, 2012; Beumer *et al*., 2010). Among the mycobacterial sequences identified, the nontuberculous mycobacteria (NTM) were relatively abundant, specifically, *Mycobacterium mucogenicum* (Fig. 4.30). Members of the NTM group are considered ubiquitous in the environment and potentially pathogenic to individuals with predisposing conditions. A notable observation was the occurrence of these mycobacteria in biofilm samples and not water samples with the exception of BRW5 from which all the identified mycobacteria were recovered. *Mycobacterium mucogenicum* was first discovered in a dialysis patient who developed septicemia via a catheter line that had been washed in infected water (Shehan and Sarma 2008). *M. mucogenicum* is associated with a wide range of clinical diseases and has been observed in complex infections of both immunocompromised and immunocompetent individuals. The bacterium is able to tolerate various disinfectants including chlorination and

extreme temperature (Reddy *et al.*, 2012). Their resistance to disinfectants can be attributed to their complex cell wall (Liu, 2012; Szewzyk *et al.*, 2000). There are also many species of mycobacteria that normally live as environmental saprophytes. These environmental mycobacteria (EM) are also opportunistic causes of disease in humans and animals. They are frequently found in drinking water distribution systemss (Liu, 2012; Vaerewijck *et al.*, 2005; September *et al.*, 2004) where they form biofilms on the surface of pipelines. Hence, eliminating mycobacteria from drinking water distribution systems is more difficult than other pathogens because of their tolerance to disinfection.

Achromobacter xylosoxidans, which was recovered in MZW1 should be considered a potential pathogen in patients with skin and soft tissue infections, especially in patients with vascular diseases or after surgery or trauma. A history of contact with water is likely to be traced during investigation of cases of *Achromobacter xylosoxidans* infection. *Dietzia maris* found in Mzima biofilms at source, has been isolated from soil and the skin and intestinal tract of carp. A case associating *D. maris* bacteremia with the presence of a catheter in an immunocompromised patient presenting with septic shock and pneumothorax has also been reported. *Micrococcus oxydans* and *paraoxydans* are known to cause Endophthalmitis, UTI, endocarditis, soft tissue infection, hypersensitivity pneumonitis, meningitis, CAPD peritonitis and bacteraemia. Several pathogenic corynebacterial species including *Corynebacterium cylea, C. macginley* and *C. tuscaniae* were recovered from Baricho source biofilm. These *Corynebacterium* species are known to cause septicaemia, peritonitis, urinary tract infections, eye infection, wound infection, endocarditis, osteomyelitis, septic arthritis, meningitis and

abscesses. Unfortunately they are also known to be multi-drug resistant and are increasingly detected in immunosuppressed patients (Tatsumi and Inui, 2012).

Pseudomonas species were recovered from all water and biofilm samples in different proportions. *Pseudomonas aeruginosa P. alcaligenes*, *P. mendocina, P. chlororaphis, P. fluorescens*, *P. mosselii, P. putida* and *P. stutzeri* are pathogenic causing bacteraemia, UTI, wound infection, abscesses, septic arthritis, conjunctivitis, endocarditis, meningitis and otitis media, most of which are nosocomial infections associated with invasive devices in debilitated patients. *Pseudomonas stutzeri* is uncommonly isolated from patients and rarely causes disease. Clinical isolates of *P. stutzeri* from a university hospital were reviewed over a 16-year period (Bisharat *et al.,* 2012). In this hospital experience, only three patients were identified with *P. stutzeri* infection, and in only one of these was the organism present as the sole isolate. A review of the literature shows that *P. stutzeri* is most frequently isolated from blood, wounds, the respiratory tract, and urine. *P. stutzeri* has also been associated with septicemia and bacteremia, as well as endocarditis on a prosthetic mitral valve, Meningitis due to *P. stutzeri* has also been described (Sunbul *et al.,* 2009). Bone and joint infections included vertebral osteomyelitis. Other reports include continuous ambulatory peritoneal dialysis (CAPD), *P. stutzeri* associated peritonitis (Ceri *et al.,* 2010), conjunctivitis and corneal infections. Cases of nosocomial transmission of *P. stutzeri* include secondary bacteremia caused by contamination of a dialysate, in hemodialysis patients (Goetz *et al.,* 1983). Patients with *P. stutzeri* infections often have serious underlying disease but generally respond to treatment with antibiotics including

the aminoglycosides, the antipseudomonal penicillins, trimethoprim—sulfamethoxazole, and the third-generation cephalosporins. *P. oryzihabitans* (*Flavimonas oryzihabitans*) found in two of the Mzima water samples causes septicaemia, eye infection, CAPD and peritonitis.

Bacillus anthracis which causes anthrax was identified in eight water sample predominantly of Mzima line, and two biofilm samples of the Baricho line. *Bacillus spp*. rarely cause focal and systemic sepsis, some of whose isolates are resistant to vancomycin. *Bacillus thuringiensis* is a biological insecticide which has caused corneal infection. It was identified in very high abundance in source water both in Mzima and Baricho. Other pathogenic bacilli were *Bacillus. cereus* and *Bacillus subtilis* which cause food poisoning, wound infection, cutaneous lesions, bacteraemia, endocarditis, eye infection and self-limiting diarrhea. *Bacillus cereus* is also notably resistant to β-lactams antibiotics and may contribute antibiotic resistant genes to other bacteria in the systems.

Enterococci are known human pathogens mostly associated with water and foodborne diarrheal diseases. The pathogenic species identified in the water distribution systems were *Enterococcus gilvus, E. gallinarum, E. mundtii, E. pallens, E. Durans, E. faecalis, E. faecium and E. avium,* all implicated in bacteraemia, abscesses, endocarditis, meningitis, UTI, peritonitis, osteomyelitis and wound infections. *Enterobacter hormaechei is a* common cause of nosocomial infection characterized by bacteraemia, respiratory tract infections and UTIs (September *et al*., 2004).

Three Clostridial species (*Clostridium spp., Clostridium peptidivorans* and *Clostridium subterminale*) were identified mostly in the Mzima water line both in the water and biofilms. Of the three, only *Clostridium subterminale* is a confirmed pathogen causing wound infections, bacteraemia and abscesses. *Escherichia coli* is associated with UTIs, bacteraemia, wound infections, meningitis, enteric infections, hemolytic, uraemic syndrome. *Escherichia fergusonii* on the other hand causes bacteraemia, wound infections and UTIs. Cases of Ampicillin-resistance have been recorded.

Pathogenic *Aeromonas* species were noticeably abundant in both water lines but distinctinctly more abundant in end point water and biofilms. These included *Aeromonas enteropelogenes, A. hydrophila, A. media, A. schubertii* and *A. veronii,* which have been implicated in wound infections, abscesses, septicaemia, meningitis, leech-bite infections, alligator-bite infections and acute diarrhea infections associated with aquatic exposure. *Aeromonas spp.* has also been found in drinking water distribution systems (Sen and Rodgers, 2004).

Bacteria from the genus *Acinetobacter* which was abundant in endpoint water and biofilms of Mzima and Baricho water lines, are a common causes of nosocomial infections among immunocompromised patients, with the most common being respiratory infections of ventilated patients (Forster and Daschner, 1998). *Acinetobacter baumannii, A. junii, A. johnsonii, A. calcoaceticus, A. schindleri, A. haemolyticus, A. radioresistens* and *A. lwoffi* have been implicated in septicaemia, urinary tract infections, wound infections, abscesses, endocarditis, meningitis and osteomyelitis.

Klebsiella oxytoca and *K. variicola* found mostly in the Mzima water samples and a few Baricho biofilm samples causes urinary tract infections, bacteraemia, wound infections and respiratory tract infections. *Agrobacterium tumefaciens* found in both water lines at source causes *e*ndocarditis, CAPD peritonitis, urinary tract infections and line sepsis. This suggests that consumption of the water may be one of the causes of the increased prevalence of respiratory diseases. Use for hygienic cleansing while bathing may also be contributing to the rise in urinary tract infections especially in immunocompromised individuals.

Bacterial genera which contain plant-associated taxa such as *Burkholderia*, *Pseudomonas*, *Lactobacillus*, or *Methylobacterium* were also identified. These bacteria can also undergo bivalent interactions that enable them to colonize both plants and humans. Depending on the genera or species of these groups, the high frequency of occurrence of members of these classes in drinking water deserves some attention as they include many ubiquitous and opportunistic bacteria, with potential public health implications. For example, using culture-dependent methods, it was shown that some *Proteobacteria*, such as *Acinetobacter*, *Sphingomonadaceae*, *Pseudomonas*, *Variovorax* or *Ralstonia*, occur in tap or bottled mineral water, and may be sources of antibiotic resistance (;Casanovas-Massana and Blanch, 2012; Falcone-Dias *et al.,* 2012; Vaz-Moreira *et al.,* , 2012 and 2011a; Tokajian *et al.,* 2005; Norton and LeChevallier, 2000). Other published studies have reported on the occurrence of several pathogens in drinking water distribution systems. Such potentially potentially pathogenic mycobacteria detected in water samples collected in France (Le Dantec *et al.,* 2002).

Opportunistic pathogens, *Mycobacterium* sp., *Legionella* spp. and *P. aeruginosa* were detected in biofilms and drinking water in Germany (Emtiazi *et al.*, 2004); *Helicobacter* spp. was also identified in biofilms (Park *et al.,* 2001).

CONCLUSION

The bacterial community in drinking water distribution systems is consistent with that of other freshwater environments such as rivers and springs where distribution system waters originate.

Sequences that made up the water distribution systems dominant bacterial phyla are *Proteobacteria*, *Firmicutes*, *Bacteriodetes*, *Actinobacteria* and *Planctomycetes*. Other predominating phyla phylotypes of interest were the Candidate divisions *TM7* and *OD1*. The bacterial composition and diversity of the drinking water distribution systems of Mombasa is not different from the typical bacterial flora of water distribution systems.

The relative abundance estimates of different members of bacterial communities in Mzima and Baricho distribution systems were not similar.

There was a higher bacterial richness and abundance in the Baricho water line than the Mzima line, which had superior bacterial evenness. Equally high bacterial diversity was evident in the Baricho water line, especially in biofilms.

The bacterial community composition of drinking water distribution systems is related to its water quality.

While water from the distribution system has acceptable pH, temperature, residual Chlorine, Nitrate and Phosphate levels, Iron Lead and Manganese levels are above acceptable limits of potability. Temperature, Iron and Nitrates appear to be the most important environmental parameters in determining the composition and diversity of bacteria in the different phases of the system.

The bacterial communities of biofilm and bulk water populations are different.

Biofilms microbiomes featured characteristically higher bacterial diversity richness and abundance than water. *Gammaproteobacteria* and *Bacilli,* and *Clostridia* of the Firmicute class appeared in both biofilms and water. Water was predominated by *Alphaproteobacteria, Betaproteobcteria* and *Nitrospirae,* while *Favobacteria, Deltaproteobcteria*, bacteria *NPL.UPA2* and Candidate division *ODI* were characteristic of biofilms. A significant fraction of unclassified bacteria biofilm samples which showed that there are still many unknown bacteria remaining to be characterized in these oligotrophic environments.

There is a significant variation in bacterial composition and diversity at the source and end point phases of the water distribution systems.

Persistent dominat bacteria populations which survived different phases of the distribution steps and appeared in the two water lines were *Proteobacteria, Firmicutes* and *Bacteroidetes.* The *Nitrospirae, Elusimicrobia, Cyanobacteria, Gemmatimonadetes, NP-UPA2,* and Candidate Divisions *TM7, OP1*1 and *OD1* were present only at source but not at endpoints and mostly in biofilms. Substantial differences were observed in bacterial community diversity and abundance at source and endpoint biofilms but not in bulk water.

The water distribution system is rich in facultative and potentially pathogenic bacteria.

Presence of potentially pathogenic bacteria in both bulk water and biofilms with 140 phylotypes of potentially pathogenic species including; *Mycobacterium, Pseudomonas, Escherichia, Shigella, Aeromonas, Enterobacter* and *Bdelovibrio* were detected.

RECOMMENDATIONS

The evident presence of pathogens in these water distribution systems indicates potential health threat for humans. This can be minimized through;

Adherence to stipulated water disinfection strategies

More frequent sampling of individual water distribution systems to identify their unique challenges of contamination.

Implementation of cross connection detection programs involving real-time water quality sensors to detect and minimize possibility of chemical or microbiological contamination.

Implementation of a leak detection system and repair program.

Adoption of DNA pyrosequencing-based bacterial identification can be a valuable tool that markedly improves bacterial pathogen identification.

The identification of sequences unrelated to common microbial indicators demonstrates the presence of yet-to-be-characterized disinfectant-resistant microbial risks in drinking water, suggesting the importance of developing additional microbial indicators and techniques for the monitoring of these microbial risks.

Areas for further research include;

Isolation of candidate phyla and to continue analysing all members of the phylum Proteobacteria prominent in Water distribution systems.

Effects of seasonality on bacterial community structure.

Functional analyses of bacterial groups and associated functional genes and their potential role in disinfectant-resistance allowing for bacteria survival mechanisms in Water distribution systems.

Further genomic DNA sequencing should be performed to find the other microbial (Archeae, viruses and eukaryotic microbes) members of the communities in oligotrophic environments.

Additional metagenomic approaches will a more comprehensive understanding of the microbial ecology of drinking water distribution systems relative to disinfection regimes. Such information is critical to the design of effective management practices and subsequently helps to safeguard human health.

REFERENCES

Abriouel H., Omar N.B., Molinos A.C., López R.L., Grande M.J., Martínez-Viedma P., Ortega E., Cañamero M.M. and **Galvez A.** (2008). Comparative analysis of genetic diversity and incidence of virulence factors and antibiotic resistance among enterococcal populations from raw fruit and vegetable foods, water and soil, and clinical samples. *International Journal of Food Microbiology,* 123: 38–49.

Ahmadian A., Ehn M., Hober S. (2006). Pyrosequencing: history, biochemistry and future. *Clinica Chimica Acta*, 363: 83 – 94.

Allen H.K., Donato J., Wang H.H., Cloud-Hansen K.A., Davies J., and Handelsman J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol* 8: 251–259

Allesen-Holm M., Barken K.B., Yang L., Klausen M., Webb J.S., Kjelleberg S., Molin S., Givskov M. and **Tolker-Nielsen T.** (2006). A characterization of DNA release in *Pseudomonas aeruginosa* cultures and biofilms. *Molecular Microbiology*, 59: 1114-1128.

Allison D.G., Ruiz B., SanJose C., Jaspe A. and **Gilbert P.** (1998). Extracellular products as mediators of the formation and detachment of *Pseudomonas fluorescens* biofilms. *Federation of European Microbiological Societies Microbiology Letters,* 167: 179-184.

Allison D.G., Evans D.J., Brown M.R.W. and **Gilbert P.** (1990). Possible involvement of the division cycle in dispersal of *Escherichia coli* from biofilms. *Journal of Bacteriology, 172*: 1667-1669.

Alpay-Karaoglu S., Ozgumus O.B., Sevim E., Kolayli F., Sevimi A. and **Yesilgil P.** (2007) Investigation of antibiotic resistance profile and TEM-type beta-lactamase gene carriage of ampicillin-resistant *Escherichia coli* strains isolated from drinking water. *Annals of microbiology*, 57: 281–288

Aminov R.I. and Mackie R.I. (2007). Evolution and ecology of antibiotic resistance genes. . *Federation of European Microbiological Societies Microbiology Letters*, **271:**147-161.

Andersen G.L., He Z., DeSantis T.Z., Brodie E.L. and **Zhou J.** (2010). The use of Microarrays in Microbial Ecology. In: Environmental Molecular Microbiology. Caister Academic Press, Norfolk: pp 87-109.

Andersson D.I., and Hughes D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* 8: 260–271.

Angles M.L., Chandy J., Kastl, G., Jegatheesan, V., Cox P. and **Fisher I.** (1999). Biofilms in drinking water: Influence of organic carbon and disinfection. *Water Research,* 26: 30-33.

APHA, AWWA and **WPCF** (2005). Standard Methods for the Examination of Water and Wastewater, 21st ed., New York.

Appenzeller B.M.R., Duval Y.B., Thomas F. and **Block J.C.** (2002). Influence of phosphate on bacterial adhesion onto iron oxyhydroxide in drinking water. *Environmental Science and Technology,* 36: 646-652.

Appenzeller B.M.R., Batté M., Mathieu L., Block J.C., Lahoussine V., Cavard J. and **Gatel D.** (2001). Effect of adding phosphate to drinking eater on bacterial growth in slightly and highly corroded pipes. *Water Research,* 35: 1100-1105.

Armon R., Arbel T., and **Green M.** (1998). A quantitative and qualitative study of biofilm disinfection on glass, metal and PVC surfaces by chlorine, bromine and bromochloro-5,5 dimethylhydantoin (BCDMH). *Water Science and Technology*, 38(12): 175-179.

Arrage A.A. and **White D.C.** (1997). Monitoring biofilm-induced persistence of *Mycobacterium* in drinking water systems using GFP fluorescence. In: Bioluminescence and Chemiluminescence: Molecular Reporting with Photons (Hastings, J.W., Kricka, L.J. and Stanley, P.E. Ed.). pp. 383-386. John Wiley and Sons, New York.

Bachmann R.T. and **Edyvean R.G.J.** (2006). AFM study of the colonisation of stainless steel by *Aquabacterium commune*. *International Biodeterioration and Biodegradation,* 58: 112-118.

Bachmann R.T. and **Edyvean R.G.J.** (2005). Biofouling: an historic and contemporary review of its causes, consequences and control in drinking water distribution systems. *Biofilms* 2: 197-227.

Batte M., B. Koudjonou P., Laurent L., Mathieu J., Coallier M. and **Prevost.** (2003). Biofilm responses to ageing and to a high phosphate load in a bench-scale drinking water system. *Water Research,* 37:1351-1361.

Beech I.B. (2004). Corrosion of technical materials in the presence of biofilms-current understanding and state-of-the art methods of study. *International Biodeterioration and Biodegradation,* 53: 177-183.

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

Bennett S.T., Barnes C., Cox A., Davies L., Brown C. Toward the 1,000 dollars human genome. Pharmacogenomics. 2005; 6:373 – 382

Bennett P.M. (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology*, S 153, 347–357.

Bent S.J., Pierson J.D., Forney L.J., Danovaro R., Luna G.M., Dell'anno A. (2007). Measuring species richness based on microbial community fingerprints: the emperor has no clothes. *Applied and Environmental Microbiology,* 73:2399-401.

Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456:53-59.

Bentley D.R. (2006). Whole-genome re-sequencing. *Current Opinion in Genetic Developement,* 16:545-552.

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

Beumer A., King D., Donohue M., Mistry J., Covert T. (2010). Detection of *Mycobacterium avium* subsp. *paratuberculosis* in drinking water and biofilms by quantitative PCR. *Applied Environmental Microbiology,* 76: 7367–7370.

Binladen J., Gilbert M.T.P., Bollback J.P., Panitz F., Bendixen C. (2007). The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE* 2: e197.

Biswas P., Lu C., and **Clark R.M.** (1993). A model for chlorine concentration decay in pipes. *Water Research*, 27: 1715-1724.

Block J.C., Sibille I., Gatel D., Reasoner D.J., Lykins B. and **Clark R.M.** (1997). Biodiversity in drinking water distribution systems: a brief review. In D. Sutcliffe (Ed.), The microbiological quality of water (pp. 63-71). London, UK: Freshwater Biological Association.

Block J.C. (1992). Biofilms in drinking water distribution systems. In: Biofilms science and technology (L. E. A. Melo, ed.), Kluwer, Dordrecht pp. 469-485.

Bokulich N. A., Subramanian S., Faith, J. J., Gevers D., Gordon J. I., Knight R. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10(1), 57–59.

Bos R., van der Mei H.C. and **Busscher H.J.** (1999). Physico-chemistry of initial microbial adhesive interactions - its mechanisms and methods for study. *Federation of European Microbiological Societies Microbiology Reviews,* 23: 179-229.

Brenner D., Krieg N., Staley J., Garrity G., Boone D. (2005). The proteobacteria, part B, the gammaproteobacteria. Bergey's manual of systematic bacteriology 2.

Brown M.L., Aldrich H.C. and **Gauthier J.J.** (1995). Relationship between glycocalyx and povidone-iodine resistance in *Pseudomonas aeruginosa* (ATCC 27853) biofilms. *Applied Environmental Microbiology,* 61: 187-193.

Bryers J.D. (2000). Biofilm formation and persistence. In J. D. Bryers (Ed.), Biofilms II: Process analysis and applications, (pp. 45-88). New York: John Wiley & Sons.

Burnham K.P. and **Overton W.S.** (1979). Robust Estimation of Population Size When Capture Probabilities Vary Among Animals. *Ecology,* 60: 927-936.

Busscher H.J., Bos R. and **Van der Mei H.C.** (1995). Initial microbial adhesion is a determinant for the strength of biofilm adhesion. *Federation of European Microbiological Societies Microbiology Letters,* 128: 229–234.

Buswell C.M. (1998). Extended survival and persistence of Campylobacter spp. in water and aquatic biofilms and their detection by immunofluorescent-antibody and -rRNA staining. *Applied Environmental Microbiology*, 64: 733-741.

Butterfield P.W., Camper A.K., Ellis B.D. and **Jones W.L.** (2002). Chlorination of model drinking water biofilm: implications for growth and organic carbon removal. *Water Research*, 36: 4391- 4405.

Byrd J.J., H.S. Xu, and R.R. Colwell. (1991). Viable but non-culturable bacteria in drinking water. *Applied Environmental Microbiology,* 57: 875–878.

Camper A.K. (2004). Involvement of humic substances in regrowth. *International Journal of Food Microbiology*, 92: 355 - 364.

Camper A.K., Brastrup K., Sandvig A., Clement J., Spencer C. and **Capuzzi A.J.** (2003). Effect of distribution system materials on bacterial regrowth. *Journal of American Water Works Association,* 95: 107-121.

Camper A.K.,. LeChevallier M.W., Broadaway S.C. and McFeters G.A. (1985). Evaluation of procedures to desorb bacteria from granular activated carbon. *Journal of Microbiological Methods,* 3:187-198

Canton R. (2009). Antibiotic resistance genes from the environment: a perspective through newly identified antibiotic resistance mechanisms in the clinical setting. *Clinical Microbiology of Infections*, 15(Suppl. 1):20-25.

Caporaso J.G., Bittinger K., Bushman F.D., DeSantis T.Z., Andersen G.L. and **Knight R**. (2010). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2): 266–267.

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

Casanovas-Massana .A. and **Blanch A.R.** (2012). Diversity of the heterotrophic microbial populations for distinguishing natural mineral waters. *International Journal of Food Microbiology,* 153: 38–44.

Cernat R., Balotescu C., Ivanescu D., Nedelcu D., Lazar V., Bucur M., Valeanu D., Tudorache R., Mitache M. and **Dragoescu M.** (2007). Mechanisms of resistance in multiple-antibiotic-resistant *Escherichia coli* strains isolated from drinking and recreational, salmaster waters. *International Journal of Antimicrobial Agents,* 29: S274

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

Chang Y.C., Puil M.L., Biggerstaff J., Randall A.A., Schulte A. and **Taylor J.S.** (2003). Direct estimation of biofilm density on different pipe material coupons using a specific DNA-probe. *Molecular and Cellular Probes,* 17: 237–243.

Chao A. and **Lee S.** (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association,* 87: 210-217.

Chao A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics,* 11: 265-270.

Characklis W.G., Turakhia, M.H. and **Zelver N.** (1990). Transport and interfacial transfer phenomena. In W. G. Characklis, & K. C. Marshall (Eds.), Biofilms. New York: John Wiley & Sons, Inc.

Charnock C. and **Kjønnø O.** (2000). Assimilable organic carbon and biodegradable dissolved organic in Norwegian raw and drinking waters. *Water Research*, 34(10): 2629-2642.

Chen V.B., Davis I.W. and **Richardson D.C.** (2009a). KING (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. *Protein Science*, 18(11): 2403–2409.

Chen Y., L.Q. Wu, R. Boden., A. Hillebrand., D. Kumaresan., H. Moussard., M. Baciu., Y. H. Lu. and **J. C. Murrell.** (2009b). Life without light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. *International Society for Microbial Ecology Journal,* 3:1093-1104.

Chen M.J., Zhang Z., and **Bott T.R.** (2005). Effects of operating conditions on the adhesive strength of *Pseudomonas fluorescens* biofilms in tubes. *Colloids and Surfaces B: Biointerfaces,* 43: 61-71.

Cloete T.E., Westaard D. and **Van Vuuren S.J.** (2003). Dynamic response of biofilm to pipe surface and fluid velocity. *Water Science and Technology,* 47: 57-59.

Cody M.L. and **Diamond J.M.** (1975). Ecology and Evolution of Communities. Belknap Press of Harvard University Press: Cambridge (Mass.).

Codony F., J. Morato. and **J. Mas.** (2005). Role of discontinuous chlorination on microbial production by drinking water biofilms. *Water Research*, 39:1896-1906.

Codony F., Morato J., Ribas F. and **Mas J.D.** (2002). Effect of chlorine, biodegradable dissolved organic carbon and suspended bacteria on biofilm development in drinking water systems. *Journal of Basic Microbiology,* 42: 311-319.

Colwell R.K. and **Coddington J.A.** (1994). Estimating Terrestrial Biodiversity through Extrapolation. Philosophical Transactions of the Royal Society B: *Biological Sciences* 345: 101–118.

Comeau A.M., Harding T., Galand P.E., Vincent W.F. and **Lovejoy C.** (2012). Vertical distribution of microbial communities in a perennially stratifi ed Arctic lake with saline, anoxic bottom waters. *Science Reports*, 2:604.

Costerton, J.W., Lewandowski Z., Caldwell D.E., Korber D.R. and **Lappin-Scott H.M.** (1995). Microbial biofilms. *Annual Review of Microbiology,* 49: 711-745.

Covert T.C., Rodgers M.R., Reyes A.L. and **Stelma G.N.** (1999). Occurrence of nontuberculous mycobacteria in environmental samples. *Applied Environmental Microbiology,* 65: 2492–2496.

Craun., Joan M., Brunkard., Jonathan S., Yoder., Virginia A., Roberts., Joe Carpenter., Tim Wade., Rebecca L., Calderon., Jacquelin M., Roberts., Michael J., Beach and Sharon L. Roy (2010) . Causes of Outbreaks Associated with Drinking Water in the United States from 1971 to 2006 *Clinical Microbiology Reviews*, Vol. 23, p. 507–528.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W. (2010). Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology,* 19: 4–20.

Critchley, M. M., R. Pasetto, and **R. J. O'Halloran.** (2004). Microbiological influences in 'blue water' copper corrosion. *Journal of Applied Microbiology,* 97: 590-597.

Critchley M. and **Fallowfield H.** (2001). The effect of distribution system bacterial biofilms on copper concentrations in drinking water. *Water Science and Technology; Water Supply,* 1: 247–252.

Dailloux M., Albert M., Laurain C., Andolfatto S., Lozniewski A., Hartemann P. and **Mathieu L.** (2003). *Mycobacterium xenopi* and drinking water biofilms. *Applied Environmental Microbiology*, 69: 6946-6948

Daniel R. (2011). Soil-Based Metagenomics. In: Handbook of Molecular Microbial Ecology II: Metagenomics in Diff erent Habitats. Wiley-Blackwell, pp 83-92.

Das A. (2012). Flippase-mediated phospholipid asymmetry promotes fast Cdc42 recycling in dynamic maintenance of cell polarity. Nat Cell Bio

De Beer D., Srinivasan R. and **Stewart P.S.** (1994). Direct measurement of chlorine penetration into biofilms during disinfection. *Applied and Environmental Microbiology,* 60: 4339-4344.

DeLeon-Rodriguez N., Lathem T.L., Rodriguez-R L.M., Barazesh J.M., Anderson B.E., Beyersdorf A.J., (2013). Microbiome of the upper troposphere: species composition and prevalence, eff ects of tropical storms, and atmospheric implications. Proceedings of the National Academy of Science USA, 110: 2575-2580.

Deng Y., He Z., Xu M., Qin Y., Van Nostrand J.D., Wu L. (2012). Elevated carbon dioxide alters the structure of soil microbial communities. *Applied Environmental Microbiology,* 78: 2991-2995.

Dinis J.M., Barton D.E., Ghadiri J., Surendar D., Reddy K., Velasquez F. (2011). In search of an uncultured human-associated *TM7* bacterium in the environment. *PLoS ONE* 6:e21280

Donlan R.M. (2002). Biofilms: microbial life on surfaces. *Emerging Infectious Diseases*, 8: 881-890.

Edgar R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19): 2460–2461.

Edgar R.C., Haas B.J., Clemente J.C., Quince C. and **Knight R.** (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16): 2194–2200.

Eduard Hoehn (1998). Hydrology, Water Resources and Ecology in Headwaters (Proceedings of the HeadWatcr'98 Conference held at Meran/Merano, Italy, April). IAHS Publ. no. 248, 1998.

Edwards R.A., Rodriguez-Brito B., Wegley L., Haynes M., Breitbart M. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BioMed Central Genomics*, 2164. 7: 57.

Eichler S., R. Christen., C. Holtje., P. Westphal., J. Botel., I. Brettar., A. Mehling. and **M. G. Hofle.** (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 Environmental Microbiology,* 72:1858-1872.

Elmar P., Christian Q., Katrin K.,l Bernhard M., Fuchs, Wolfgang L., Jörg P. and **Frank O. G.** (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. PMCID: PMC2175337. *Nucleic Acids Research,* 35(21): 7188–7196.

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

Ercolini, D. (2004). PCR-DGGE fingerprinting: Novel strategies for detection of microbes in food. *Journal of Microbiological Methods*, 56: 297-314.

Faith D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10.

Falcone-Dias M., Vaz-Moreira I. and **Manaia C.M.** (2012). Bottled mineral water as a potential source of antibiotic resistant bacteria. *Water Research*, 46: 3612–3622.

Falkinham J.O. III., Norton C.D. and **LeChevallier M.W.** (2001). Factors influencing numbers of *Mycobacterium avium, Mycobacterium intracellulare*, and other

Mycobacteria in drinking water distribution systems. *Applied Environmental Microbiology*, 67: 1225-1231.

Feazel L.M., Baumgartner L.K., Peterson K.L., Frank D.N., Harris J.K. and **Pace N.R.** (2009). Opportunistic pathogens enriched in showerhead biofilms. Proceedings of the National Academy of Science USA, 106: 16393 – 16399.

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

Figueras M.J. and **Borrego J.J.** (2010). New perspectives in monitoring drinking water microbial quality. *International Journal of Environmental Research and Public Health,* 7: 4179–4202

Freeman R., H. Geier., K.M. Weigel J. Do., T.E. Ford. and **G.A. Cangelosi.** (2006). Roles for cell wall glycopeptidolipid in surface adherence and planktonic dispersal of *Mycobacterium avium. Applied Environmental Microbiology,* 72: 7554-7558.

Forster D. and **Daschner F.** (1998). Acinetobacter species as nosocomial pathogens. *European Journal of Clinical Microbiology and Infectious Diseases,* 17: 73–77.

Fouts D.E., Szpakowski S., Purushe J., Torralba M., Waterman R.C., MacNeil M.D. (2012). Next generation sequencing to defi ne prokaryotic and fungal diversity in the bovine rumen. *PLoS One*, 7:e48289.

Gagnon G.A., Rand J.L., O'Leary K.C., Rygel A.C., Chauret C. and **Andrews R.C.** (2005). Disinfectant efficacy of chlorite and chlorine dioxide in drinking water biofilms. *Water Research*, 39: 1809-1817.

Geldreich E.E. (1996). Microbial Quality of Water Supply in Distribution Systems.1stEdn. New York: CRC Press.

Gentry R.W., J. McCarthy., A. Layton., L. McKay., D. Williams., S.R. Koirala. and **G.S. Sayler.** (2006). *Escherichia coli* loading at or near baseflow in a mixed-use watershed. *Journal of Environmental Quality*, 35: 2244–2249.

Gihring T.M., Green S.J. and **Schadt C.W.** (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology*, 14: 285–290.

Gilbert P., Evans D.J. and **Brown M.R.W.** (1993). Formation and dispersal of bacterial biofilms *in vivo* and *in situ*. *Journal of Appllied Bacteriology,* 74: 67S–78S.

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

Green, D. H., Shenoy, D. M., Hart, M. C., and **Hatton, A. D.** (2011). Coupling of dimethylsulfide oxidation to biomass production by a marine flavobacterium. *Applied Environmental Microbiol*ogy, 77, 3137–3140

Griffiths R.I., Thomson B.C., James P., Bell T., Bailey M. and **Whiteley A.S.** (2011). The bacterial biogeography of British soils. *Environmental Microbiology*, 13: 1642–1654.

Grobe S., J. Wingender. and **H.C. Flemming.** (2001). Capability of mucoid *Pseudomonas aeruginosa* to survive in chlorinated water. *International Journal of Hygiene and Environmental Health,* 204: 139-142.

Gu B., Schmitt J., Chen C., Liang L. and **McCarthy J.F.** (1994). Adsorption and desorption of natural organic matter on iron oxide: mechanisms and models. *Environmental Science and Technology*, 28: 38–46.

Hamady M., Walker J.J., Harris J.K, Gold N.J., Knight R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* 5:235–237

Hansen S.K., Rainey P.B., Haagensen J.A. and **Molin S.** (2007). Evolution of species interactions in a biofilm community. *Nature*, 445: 533–536.

Hanski I., von Hertzen L., Fyhrquist N., Koskinen K., Torppa K., Laatikainen T. (2012). Environmental biodiversity, human microbiota, and allergy are interrelated. Proceedings of the National Academy of Sciences.

Havelaar A.H., Versteegh J.F.M. and **During M.** (1990). The presence of Aeromonas in drinking-water supplies in the Netherlands. *Zentralbl.Hyg.*190: 236-56.

Heltshe J.F. and **Forrester N.E.** (1983). Estimating Species Richness Using the Jackknife Procedure. *Biometrics*, 39:1-11.

Hem L.J. and **Skjevrak I.** (2002). Potential water quality deterioration of drinking water caused by leakage of organic compounds from materials in contact with the water. In Copenhagen: 20th NoDig Conference. Biofilms in Drinking Water 29.

Henne K., Kahlisch L., Brettar I. and Hofle M.G. (2012). Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and References 63 bulk water of a citywide network in Germany. *Appllied Environmenta Microbiology,* 78 (10):3530-3538.

Hijnen W.A.M., Beerendonk E.F. and **Medema G.J.** (2006). Inactivation credit of UV radiation for viruses, bacteria and protozoan (oo)cysts in water: a review. *Water Research,* 40: 3-22.

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

Höfle M., Kirchman D., Christen R. and **Brettar I.** (2008). Molecular diversity of bacterioplankton: link to a predictive biogeochemistry of pelagic ecosystems. *Aquatic Microbial Ecology,* 53: 39–58.

Hofreuter D., Tsai J., Watson R.O., Novik V., Altman B., Benitez M., Clark C., Perbost C., Jarvie T. (2006). Unique features of a highly pathogenic *Campylobacter jejuni* strain. *Infection Immunology*, 74 : 4694 – 4707.

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

Hong P., Hwang C., Ling F., Andersen G.L., LeChevallier M.W. (2010). Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Applied Environmental Microbiology*, 76: 5631–5635.

Horn H., Wasche S. and **Hempel D.C.** (2002). Simulation of biofilm growth, substrate conversion and mass transfer under different hydrodynamic conditions. *Water Science and Technology,* 46: 249-252.

Hortal J., Borges P.A. and **Gaspar C.** (2006). Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology,* 75: 274-287.

Hoyle B.D., Alcantara J. and **W C.J.** (1992). *Pseudomonas aeruginosa* biofilm as a diffusion barrier to piperacillin. *AntimicrobialAgents Chemotherapy.*

Hu F.B, Rimm E., Smith-Warner .SA., Feskanich D., Stampfer M.J., Ascherio A., Simpson L. and Willet W.C. (1999): Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *American Journalof Clinical Nutrition,* 69: 243–249.

Hu J.Y., Yu B., Feng Y.Y., Tan X.L., Ong S.L., Ng W.J. and **Hoe W.C.** (2005). Investigation into biofilms in a local drinking water distribution system. *Biofilms,* 2: 19-25.

Huang S., Yang F., Zeng X., Chen J., Li R., Wen T. (2011). Preliminary characterization of the oral microbiota of Chinese adults with and without gingivitis. *BioMed Central Oral Health,* 11:33-6831-11-33.

Huber J.A. Mark Welch D.B, Morrison H.G, Huse S.M, Neal PR, Butterfield DA and **Sogin M.L** (2007). Microbial population structures in the deep marine biosphere. *Science*, 318: 97–100.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology,* 180**:** 4765–4774

Hugenholtz P. and **Tyson G.W.** (2008). Microbiology: Metagenomics. *Nature,* 455: 481-483.

Huirong L., Shuting Z., Shenghua Z., Wenfang L. and **Xin Y.** (2015).The function of advanced treatment process in a drinking water treatment plant with organic matter-polluted source water. *Environmental Science and Pollution Research,* 015-5116

Huse S.M., Dethlefsen L., Huber J.A., Welch D.M., Relman D.A., Sogin M.L. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics*. ;4 e1000255.

Hulcr J., Latimer A.M., Henley J.B., Rountree N.R., Fierer N. and **Lucky A.** (2012). A Jungle in There: Bacteria in Belly Buttons are highly diverse, but predictable. *PLoS ONE* 7: e47712.

Huse S.M., Welch D.M., Morrison H.G. and **Sogin M.L.** (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology,* 12:1889-1898.

Iversen A., Kühn I., Rahman M., Franklin A., Burman L.G., Olsson- Liljequist B., Torell E. and **Möllby R.** (2004). Evidence for transmission between humans and the environment of a nosocomial strain of *Enterococcus faecium*. *Environmental Microbiology,* 6: 55–59

Jenkins, O., D. Byrom. and **D. Jones.** (1987). *Methylophilus*—a new genus of methanol-utilizing bacteria. *International Journal of Systematic Bacteriology*, 37: 446-448.

Juottonen H. Tuittila E.S., Juutinen S., Fritze H. and **Yrjala K.** (2008). Seasonality of rDNA- and rRNA-derived archaeal communities and methanogenic potential in aboreal mine. *International Society for Microbial Ecology Journal,* 2:1157-1168.

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

Kalmbach S., W. Manz., and U. Szewzyk. (1997). Dynamics of biofilm formation in drinking water: phylogenetic affiliation and metabolic potential of single cells assessed by formazan reduction and in situ hybridization. *Federation of European Microbiological Societies for Microbiol Ecology,* 22: 265–280.

Kalmbach S., Manz W., Bendinger B. and **Szewzyk U.** (2000). *In situ* probing reveals *Aquabacterium* commune as a widespread and highly abundant bacterial species in drinking water biofilms. *Water Research,* 34: 575-581.

Kalmbach, S., Manz, W. and Szewzyk. U. (1997). Dynamics of biofilm formation in drinking water: phylogenetic affiliation and metabolic potential of single cells assessed by formazan reduction and in situ hybridization. *Federation of European Microbiological Societies for Microbiol Ecology,* 22:265-279.

Kalmbach, S., Manz, W. and Szewzyk, U. (1997). Isolation of new bacterial species from drinking water biofilms and proof of their in situ dominance with highly specific 16S rRNA probes. *Appllied Environmental Microbiology,* 63:4164-4170

Kalyuhznaya M.G., W. Martens-Habbena., T. Wang., M. Hackett., S.M. Stolyar., D.A. Stahl., M.E. Lidstrom. and **L. Chistoserdova.** (2009). *Methylophilaceae* link methanol oxidation to denitrification in freshwater lake sediment as suggested by stable isotope probing and pure culture analysis. *Environmental Microbiology Reports*, 1:385-392.

Kasschau K.D, Fahlgren N., Chapman E.J., Sullivan C.M., Cumbie J.S., Givan S.A., Carrington J.C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biology,* 5:e57.

Keevil C.W. and **Walker J.T.** (1992). Nomarski DIC microscopy and image analysis of biofilms. *Binary Computing Microbiology,* 4: 93-95.

Keinänen-Toivola M.M., Revetta R.P., and Santo Domingo J.W. (2006). Identification of active bacterial communities in a model drinking water biofilm system using 16S rRNA-based clone libraries. *Federation of European Microbiological Societies for Microbiology Letters*. 257, 182–188

Kemper N., (2008). Veterinary antibiotics in the aquatic and terrestrial environment. Ecol Indic 8:1–13 Kim SH, Wei CI, Tzou YM, An HJ (2005) Multidrug-resistant *Klebsiella pneumoniae* isolated from farm environments and retail products in Oklahoma. *Journal of food protection,* 68: 2022–2029

Kerr C.J., Osborn K.S., Robson G.D. and **Handley P.S.** (1999). The relationship between pipe material and biofilm formation in a laboratory model system. *Journal of Applied Microbiology,* 85: 29S – 38S.

Kerr C.J., Osborn,K.S., Rickard A.H., Robson G. D. and Handley, P.S. (2003). Biofilms in water distribution systems. In Water and Wastewater Engineering, pp. 757–776. Edited by D. Mara & N. J. Horan. London: Academic Press

Kirchman D.L (2008). Introduction and overview. In: Microbial Ecology of the Oceans. Wiley-Blackwell: New Jersey: pp 1-26.

Kokare C.R., Chakraborty S., Khopade A.N. and **Mahadik K.R.** (2009). Biofilm: importance and applications. *Indian Journal of Biotechnology,* 8: 159-168.

Koksal F., Oguzkurt N., Samasti M. and **Altas K.** (2007). Prevalence and antimicrobial resistance patterns of Aeromonas strains isolated from drinking water samples in Istanbul, Turkey. *Chemotherapy,* 53: 30–35

Kormas K.A., C. Neofitou., M. Pachiadaki. and **E. Koufostathi.** (2010). Changes of the bacterial assemblages throughout an urban drinking water distribution system. *Environmental Monitoring and Assessment*, 165:27-38.

Koskinen R., T. Ali-Vehmas., P. Kampfer., M. Laurikkala., I. Tsitko., E. Kostyal., F. Atroshi., and **M. Salkinoja-Salonen.** (2000). Characterization of Sphingomonas isolates from Finnish and Swedish drinking water distribution systems. *Journal of Applied Microbiology,* 89: 687-696

Kunimoto D.Y., Peppler M.S., Talbot J., Phillips P. and **Shafran S.D.** (2003). Analysis of *Mycobacterium avium* complex isolates from blood samples of AIDS patients by pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, 41: 498-499.

Kunin V., Engelbrektson A., Ochman H. and **Hugenholtz P.** (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artifi cial infl ation of diversity estimates. *Environmental Microbiology,* 12:118-123.

Langmark J., Storey M.V., Ashbolt N.J. and **Stenstrom T-A.** (2005). Accumulation and fate of microorganisms and microspheres in biofilms formed in a pilot-scale water distribution system. *Applied Environmental Microbiology*, 71: 706-712.

LeChevallier M.W., Welch N.J. and **Smith D.B.** (1996). Full-scale studies of factors related to coliform regrowth in drinking water. *Applied Environmental Microbiology*, 62: 2201-2211.

LeChevallier M.W., Lowry C.D., Lee R.G. and **Gibbon D.L.** (1993). Examining the Relationship between Iron Corrosion and the Disinfection of Biofilm Bacteria. *Journal American Water Works Association,* 85: 111-123.

LeChevallier M.W., Cawthon C.D. and **Lee R.G.** (1988). Factors promoting the survival of bacteria in chlorinated water supplies. *Applied and Environmental Microbiology,* 54: 649-654.

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

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

Lee S.H. and **Kim S.J.** (2002). Detection of infectious enteroviruses and adenoviruses in tap water in urban areas in Korea. *Water Research,* 36: 248-256.

Lee S.H., O'Connor J.T. and **Banerji S.** (1980). Biologically mediated corrosion and its effects on water quality in distribution systems. *Journal American. Water Works Association,* 34: 203-221

Le Dantec C., Duguet J.P., Montiel A., Dumoutier N., Dubrou S. and **Vincent V.** (2002). Chlorine disinfection of atypical *Mycobacteria* isolated from a water distribution system. *Applied and Environmental Microbiology,* 68: 1025-1032.

Lee D-G., Lee J-H. and **Kim S-J.** (2005). Diversity and dynamics of bacterial species in a biofilm at the end of the Seoul water distribution system. *World Journal Microbiol Biotechnol*, 21:155-162.

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

Lehtola M.J., Ilkka T., TiiaLampola M., Hirvonen A., Vartiainen T. and **Martikainen P.J.** (2005). Pipeline materials modify the effectiveness of disinfectants in drinking water distribution systems. *Water Research*, 39: 1962-1971.

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

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

Le Roy T., Llopis M., Lepage P., Bruneau A., Rabot S. and **Bevilacqua C.** (2012). Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. *Gut*, 62(12):1787-94.

Li, R. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20**, 265-272

Lipponen M.T., Suutari M.H. and **Martikainen P.J.** (2002). Occurrence of nitrifying bacteria and nitrification in finish drinking water distribution systems. *Water Research*, 36: 4319–4329.

Liu R. (2012). Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. *Science of the Total Environment*, 435-436, 124–131

Liu Y., Zhang W., Sileika T., Warta R. and **Cianciotto N.P.** (2011). Disinfection of bacterial biofilms in pilot-scale cooling tower systems. *Biofouling,* 27: 393–402.

Liu Y., Lin Y. M., Yang S.F. and **Tay J.H.** (2003). A balanced model for biofilms developed at different growth and detachment forces. *Process Biochemistry,* 38: 1761-1765.

Liu Y. and **Tay J.H.** (2001). Detachment forces and their influence on the structure and metabolic behaviour of biofilms. *World Journal of Microbiol Biotechnology,* 17 (2): 111-117.

Lozupone C. and **Knight R.** (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12): 8228–8235.

Lund, V. and **Ormerod, K.** (1995). The influence of disinfection processes on biofilm formation in water distribution systems. *Water Research,* 29(4): 1013-1021.

Magurran, A.E. (2004). *Measuring biological diversity.* Blackwell Publishing; Malden, Massachusetts, USA. pp 264.

Manuel C.M., Nunes O.C. and **Melo L.F.** (2010). Unsteady state flow and stagnation in distribution systems affect the biological stability of drinking water. *Biofouling* 26: 129–139.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S, Bemben L.A., Berka J., Braverman M.S., Chen Y.J. (2005). Genome sequencing in micro-fabricated high-density picolitre reactors. *Nature,* 437:376 – 380.

Manuel C.M., Nunes O.C., and Melo L.F. (2007). Dynamics of drinking water biofilm in flow/non-flow conditions. *Water Res*earch 41, 551–562

Mao S., Zhang R., Wang D. and **Zhu W.** (2012). Th e diversity of the fecal bacterial community and its relationship with the concentration of volatile fatty acids in the feces during subacute rumen acidosis in dairy cows. *BioMed Central Vetrinary Research.,* 8: 237-6148-8-237.

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

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S. and **Bemben L.A.** (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437: 376-380.

Martiny A.C., Albrechtsen H.J., Arvin E. and **Molin S.** (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.

Martiny A.C., T.M. Jorgensen., H.J. Albrechtsen., E. Arvin. and **S. Molin.** (2003). Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system. *Applied and Environmental Microbiology,* 69: 6899-6907

Mathieu L., Bouteleux C., Fass S., Angel E. and **Block J.C**. (2009). Reversible shift in the alpha-, beta-, and gammaproteobacteria populations of drinking water biofilms during discontinuous chlorination. *Water Research,* 43: 3375–3386

Maxam A.M. and **Gilbert W.** (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences USA, 74: 560-564.

McDonald D., Clemente J.C., Kuczynski J., Rideout J.R., Stombaugh J. and **Wendel D.** (2012). The Biological Observation Matrix (BIOM) format or: How I learned to stop worrying and love the ome-ome. *Gigascience*, 1(1): 7.

McKeon D.M., Calabrese J.P. and **Bissonnette G.K.** (1995). Antibiotic resistant gram-negative bacteria in rural groundwater supplies. *Water Research,* 29: 1902–1908.

McMurdie P.J. and **Holmes S.** (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4): e61217.

Melo L.F. and **Vieira M.J.** (1999). Physical stability and biological activity of biofilms under turbulent flow and low substrate concentration. *Bioprocess Engineering,* 20: 363-368.

Messi P., Guerrieri E. and **Bondi M.** (2002). Survival of an Aeromonashydrophila in an artificial mineral water microcosm. *Water Research*, 36: 3410-3415.

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

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

Miettinen I.T., Vartiainen T. and **Martikainen P.J.** (1997). Phosphorus and bacterial growth in drinking-water. *Applied and Environmental Microbiology,* 63: 3242-3245.

Modin O., K. Fukushi. and K. Yamamoto. (2007). Denitrification with methane as external carbon source. *Water Research*, 41: 2726-2738.

Molin S. and **Tolker-Nielsen T.** (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Current Opinion in Biotechnology*, 14: 255- 261.

Nagy L.A. and **Olson B.H.** (1985). Occurrence and significance of bacteria, fungi, and yeasts associated with distribution pipe surfaces. In: Proceedings of the American Water Works Association, Water Quality Technical Conferenc. pp. 213.American Water Works Association, Denver, Colorado.

Nam Y.D., Park S.L. and **Lim S.I.** (2012). Microbial composition of the Korean traditional food "kochujang" analyzed by a massive sequencing technique. *Journal of Food Science*, 77: M250-6.

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

Nieminen T.T., Valitalo H., Sade E., Paloranta A., Koskinen K. and **Bjorkroth J.** (2012). The effect of marination on lactic acid bacteria communities in raw broiler fillet strips. *Frontiers in Microbiology*, 3: 376.

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

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

Norton C.D., LeChevallier M.W., Falkinham I. and **Joseph O.** (2004). Survival of *Mycobacterium avium* in a model distribution system. *Water Research*, 38: 1457-1466.

Nyyssonen M., Hultman J., Ahonen L., Kukkonen I., Paulin L. and **Laine P.,** Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *International Society for Microbial Ecology Journal.*

Obst U., Schwartz T. and **Volkmann H.** (2006). Antibiotic resistant pathogenic bacteria and their resistance genes in bacterial biofilms. *International Journal of Artificial Organs,* 29: 387–394

Oh J.D, Kling-Backhed H., Giannakis M., Xu J., Fulton R.S., Fulton L.A., Cordum H.S., Wang C., Elliott G. (2006). The complete genome sequence of a chronic atrophic gastritis Helicobacter pylori strain: evolution during disease progression. Proc. Natl Acad. Sci. USA. ; 103:9999–10004.

Ollos P.J. (1998). Effects of Drinking Water Biodegradability and Disinfectant Residual on Bacterial Regrowth, University of Waterloo, Waterloo, Canada.

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

Oros-Sichler M., Costa R., Heuer H. and **Smalla K.** (2007). Molecular Fingerprinting Techniques to Analyze Soil Microbail Communities. In: Modern Soil Microbiology. CRC Press, pp 355-386.

Owen D.M., Amy G.L., Chowdhury Z.K., Paode R., McCoy G. and **Viscosil K.** (1995). NOM characterization and treatability. *Journal American Water Works Association,* 87: 46–63.

Ozgumus O.B., Celik-Sevim E., Alpay-Karaoglu S., Sandalli C. and **Sevim A.** (2007). Molecular characterization of antibiotic resistant *Escherichia coli* strains isolated from tap and spring waters in a coastal region in turkey. *Journal Microbiology,* 45: 379–387

Paris T., Skali-Lami S. and **Block J.C.** (2007). Effect of wall shear rate on biofilm deposition and grazing in drinking water flow chambers. *Biotechnology and Bioengineering,* 97: 1550-1561.

Park S.K., Kim Y.K. and **Choi S.C.** (2008). Response of microbial growth to orthophosphate and organic carbon influx in copper and plastic based plumbing water systems. *Chemosphere,* 72: 1027-1034.

Park S.R., Mackay W.G. and **Reid D.C.,** (2001). Helicobacter sp. recovered from drinking water biofilm sampled from a water distribution system. *Water Research,* 35:1624-1626.

Parsek M.R. and **Singh P.K.** (2003). Bacterial biofilms: an emerging link to disease pathogenesis. *Annual Reviewof Microbiology,* 57: 677-701.

Payment, P. (1999). Poor efficacy of residual chlorine disinfectant in drinking water to inactivate waterborne pathogens in distribution systems. *Canadian Journal Microbiology*, 45: 709–715

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends in Microbiology,* 14: 257–263.

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

Picioreanu C., Van Loosdrecht M.C.M. and **Heijnen J.J.** (2001). Two dimensional model of biofilm detachment caused by internal stress from liquid flow. *Biotechnology and Bioengineering,* 72: 205-218.

Poindexter J. S., Kanan P. Pujara and James T. Staley (2000). : In Situ Reproductive Rate of Freshwater *Caulobacter spp... Applied Environmental Microbiology,* 66(9):4105-4111.

Poitelon J.B., (2010). Variations of bacterial 16S rDNA phylotypes prior to and after chlorination for drinking water production from two surface water treatment plants. *Journal Indian Microbiology and Biotechnology,* 37: 117–128

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

Pozos N., Scow K., Wuertz S. and **Darby J.** (2004). UV disinfection in a model distribution system: biofilm growth and microbial community. *Water Research*, 38: 3083-3091.

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

Prabhu D.I.G., Pandian R.S. and **Vasan P.T.** (2007). Pathogenicity, antibiotic susceptibility and genetic similarity of environmental and clinical isolates of Vibrio cholerae. *Indian Journal of Experimental Biology,* 45: 817–823

Pryor M., Springthorpe S., Riffard S., Brooks T., Huo Y., (2004). Investigation of opportunistic pathogens in municipal drinking water under different supply and treatment regimes. *Water Science and Technology*, 50: 83–90.

Qiu Q., R. Conrad. and **Y. Lu.** (2009). Cross-feeding of methane carbon among bacteria on rice roots revealed by DNA-stable isotope probing. *Environmental Microbiology Reports,* 1: 355-361.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596.

Quince C., Lanzen A., Davenport R.J. and **Turnbaugh P.J**. (2011). Removing noise from pyrosequenced amplicons. *BioMed Central Bioinformatics*, 12: 38.

Ram S., Poornima V., and Rishi S. (2007). Contamination of Potable Water Distribution Systems by Multiantimicrobial-Resistant Enterohemorrhagic *Escherichia coli.*

Rastogi G. and **Sani R.** (2011). Molecular Techniques to Assess Microbial Community Structure, Function, and Dynamics in the Environment. *Springer* New York, pp 29-57.

Reddy Y., Balasubramaniam L., Mathew M. and Abraham G. (2012). *Mycobacterium mucogenicum* Peritonitis in a Continuous Ambulatory Peritoneal Dialysis Patient. *Peritoneal Dialysis International*: 226-227.

Regan, J.M., Harrington, G.W., Baribeau, H. De Leon, R. and Noguera D.R. (2003). Diversity of nitrifying bacteria in full-scale chloraminated distribution systems. *Water Research.* 37:197-205

Regan J.M., Harrington G.W. and **Noguera D.R.** (2002). Ammonia- and nitrite-oxidizing bacterial communities in a pilot-scale chloraminated drinking water distribution system. *Applied Environmental Microbiology,* 68: 73–81.

Revetta R.P., Pemberton A., Lamendella R., Iker B. and Santo Domingo J.W. (2010). Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water Research,* 44: 1353–1360.

Revetta R.P., Matlib R.S. and **Santo Domingo J.W.** (2011). 16S rRNA gene sequence analysis of drinking water using RNA and DNA extracts as targets for clone library development. *Current Opinions in Microbiology,* 63: 50–59

Revetta R.P., Gomez-Alvarez V., Gerke T.L., Curioso C., Santo Domingo J.W. and Ashbolt N.J. (2013). Establishment and early succession of bacterial communities in monochloramine-treated drinking water biofilms. *Federation of European Microbiological Societies for Microbiol Ecology*, 12170:1574-6941.

Ribas F., J. Perramon., A. Terradillos., J. Frias. and **F. Lucena.** (2000). The *Pseudomonas* group as an indicator of potential regrowth in water distribution systems. *Journal of Applied Microbiology,* 88: 704-710.

Rickard A.H., Stead A.T., O'May G.A., Lindsay S., Banner M., Handley P.S. and **Gilbert P.** (2005). *Adhaeribacter aquaticus* gen. *nov.*, sp. *nov.*, a Gram-negative isolate from a potable water biofilm. *International journal of systematic and evolutionary microbiology*, 55: 821-829.

Ridgway H. and **Olson B.** (1982). Chlorine resistance patterns of bacteria from two drinking water distribution systems. *Applied and Environmental Microbiology,* 44: 972–987.

Roberts M.E. and **Stewart P.S.** (2005): Modelling protection from antimicrobial agents in biofilms through the formation of persister cells. *Microbiology*, 151: 75-80.

Rodríguez C., Lang L., Wang A., Altendorf K., García F. and **Lipski A.** (2006). Lettuce for human consumption collected in Costa Rica contains complex communities of culturable oxytetracycline-and gentamicin-resistant bacteria. *Applied Environmental Microbiology,* 72: 5870– 5876

Rogers J., Dowsett A.B., Dennis P.J., Lee J.V. and **Keevil C.W.** (1994). Influence of plumbing materials on biofilm formation and growth of *Legionella pneumophila* in potable water systems. *Applied and Environmental Microbiology*, 60: 1842-1851.

Rosenberg E. and **Ron E.Z.** (1999). High- and low-molecular-mass microbial surfactants. *Applied Microbiology and Biotechnology,* 52: 154-162.

Rothberg J.M., Wolfgang H., Todd M., Rearick J.S., William M., Mel Davey., John H., Leamon, Kim J., Mark J., Milgrew., Matthew E., Jeremy H., Jan F.S., David M., Jason W., Myers., John F., Davidson., Annika B., John R. Nobile., Bernard P.P., David L., Travis A., Clark., Martin H., Jeffrey T., Branciforte, Isaac B., Stoner., Simon E., Cawley., Michael L. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352

Routledge R.D. (1984). Estimating Ecological Components of Diversity. *Oikos,* 42:23-29.

Routledge R.D. (1977). On Whittaker's Components of Diversity. *Ecology,* 58: 1120-1127.

Ruby J.G., Jan C., Player C., Axtell M.J., Lee W., Nusbaum C., Ge H., Bartel D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell,* 127:1193–1207.

Ryan P.G., Moore C.J., van Franeker J.A., Moloney C.L. (2009). Monitoring the abundance of plastic debris in the marine environment. Philosophical Transanctions of the Royal Society, B 364: 1999–2012.

Sambrook K.J., Fritsch E.F. and **Maniatis T.** (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger F., Nicklen S. and **Coulson A.R.** (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA, 74: 5463-5467.

Sanger F. and **Coulson A.R.** (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology,* 94: 441-448.

Santo Domingo J.W., Meckes M.C., Simpson J.M., Sloss B., and **Reasoner D.J.** (2003). Molecular characterization of bacteria inhabiting a water distribution system simulator. *Water Science and Technolology,* 47: 149-154.

Sarmah A.K., Meyer M.T. and **Boxall A.B.A.,** (2006). A global perspective onthe use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere,* 65:725–759

Sathasivan A. and **Ohgaki, S.** (1999). Application of new bacterial regrowth potential method for water distribution system - a clear evidence of phosphorus limitation. *Water Research,* 33: 137-144.

Sauer K., Cullen M.C., Rickard A.H., Zeef L.A.H., Davies D.G. and **Gilbert P.** (2004). Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *Journal of Bacteriology,* 186: 7312-7326.

Schloss P.D. and **Handelsman, J**. (2006). Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Applied and Environmental Microbiology*, 72(10): 6773–6779.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M. and **Hollister E.B.** (2009). Introducing mothur: Open-source, platform-independent, community supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23): 7537–7541.

Schmidt T.M., DeLong E.F. and **Pace N.R.** (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology,* 173: 4371–4378.

Schwartz T., Hoffmann S. and **Obst U.** (2003). Formation of natural biofilms during chlorine dioxide and u.v. disinfection in a public drinking water distribution system. *Journalof Applied Microbiology*, 95: 591-601.

Schwartz T., Kohnen W., Jansen B. and **Obst U.** (2003). Detection of antibiotic-resistant bacteria and their resistance genes in waste water, surface water, and drinking water biofilms. *Federation of European Microbiological Societies Microbiol Ecology*, 43: 325-335.

Schwartz T., Hoffmann S. and **Obst U.** (1998). Formation and bacterial composition of young, natural biofilms obtained from public bank-filtered drinking water systems. *Water Research,* 32: 2787-2797.

Schwarz D.S., Hutvágner G., Du T., Xu Z., Aronin N., and Zamore P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell,* 115:199-208.

Szewzyk U., W. Manz, R. Amann, K.H. Schleifer and T.A. Stenstrom. (1994). Growth and in situ detection of a pathogenic *Escherichia coli* in bio-films of a heterotrophic water-bacterium by use of 16S- and 23S-rRNA- directed fluorescent oligonucleotide probes. *Federation of European Microbiological Societies Microbiol Ecology.* 13: 169 – 176

Sen K. and **Rodgers M.** (2004). Distribution of six virulence factors in *Aeromonas* species isolated from US drinking water utilities: a PCR identification. *Journal of Applied Microbiology*, 97: 1077-1086.

September S.M., Brözel V.S. and **Venter S.N.** (2004). Diversity of nontuberculoid Mycobacterium species in biofilms of urban and semiurban drinking water distribution systems. *Applied and Environmental Microbiology*, 70: 7571–7573.

Servais P., Anzil A., Gatel D. and **Cavard J.** (2004). Biofilms in the Parisian suburbs drinking water distribution system. *Journal Water Supply Research and Technology - Aqua*, 53: 313-324.

Servais P., Laurent P. and **Randon G.** (1995). Comparison of the Bacterial Dynamics in Various French Distribution-Systems. *Journal of Water Supply Research and Technology-Aqua,* 44: 10-17.

Shade A. and **Handelsman J.** (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology,* 14: 4–12.

Shehan J.M. and **Sarma D.P.** (2008). Mycobacterium mucogenicum: report of a skin infection associated with etanercept. *Dermatology Online Journal,* 14:5

Shendure J., Porreca G., Reppas N., Lin X., McCutcheon J. and **Rosenbaum A.** (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science (New York, N.Y.) 309:1728-1732.

Shirey J.J. and **Bissonnette G.K.** (1991). Detection and identification of groundwater bacteria capable of escaping entrapment on 0.45-micron-pore-size membrane filters. *Applied Environmental Microbiology,* 57: 2251–2254.

Shokralla S., Spall J.L., Gibson J.F. and **Hajibabaei M.** (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology,* 21: 1794–1805.

Siam R., Mustafa G.A., Sharaf H., Moustafa A., Ramadan A.R., Antunes A. (2012). Unique prokaryotic consortia in geochemically distinct sediments from Red Sea Atlantis II and discovery deep brine pools. *PLoS One,* 7:e42872.

Simões L.C., Simões M. and **Vieira M.J.** (2010). Influence of the diversity of bacterial isolates from drinking water on resistance of biofilms to disinfection. *Applied and Environmental Microbiology,* 76: 6673-6679.

Simoes L.C., M. Simoes and **M.J. Vieira.** (2007). Biofilm interactions between distinct bacterial genera isolated from drinking water. *Applied and Environmental Microbiology*, 73**:** 6192-6200.

Simões L.C., Azevedo N., Pacheco A., Keevil C.W. and **Vieira M.J.** (2006). Drinking water biofilm assessment of total and culturable bacteria under different operating conditions. *Biofouling,* 22: 91-99.

Singleton D.R., Furlong M.A., Rathbun S.L. and **Whitman W.B.** (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Applied Environmental Microbiology,* 67:4374-4376.

Smith C.A., Phiefer C.B., MacNaughton S.J., Peacock A., Burkhalter R.S., Kirkegaard R. and **White D.C.** (2000). Quantitative lipid biomarker detection of unculturable microbes and chlorine exposure in water distribution system biofims. *Water Research,* 34(10): 2683-2688.

Smith E.P. and **Van Belle G.** (1984). Nonparametric Estimation of Species Richness. *Biometrics,* 40:119-129.

Soini S.M., Koskinen K.T., Vilenius M.J. and **Puhakka J.A.** (2002). Effects of fluid flow velocity and quality on planktonic and sessile microbial growth in water hydraulic systems. *Water Research,* 36: 3812-3820.

Sorokin D. Y. (1996). Oxidation of sulfide and elemental sulfur to tetrathionate by chemoorganoheterotrophic bacteria. *Microbiology*, 65, 1–5

Srinivasan S., Harrington, G. W., Xagoraraki, I. and **Goel, R.** (2008). Factors affecting bulk to total bacteria ratio in drinking water distribution systems. *Water Research,* 42: 3393–3404

Srinivasan V., Nam H.M., Nguyen L.T., Tamilselvam B., Murinda S.E. and **Oliver S.P.** (2005). Prevalence of antimicrobial resistance genes in *Listeria monocytogenes* isolated from dairy farms. *Foodborne Pathogens and Diseases,* 2: 201–211

Stahl P.L. and **Lundeberg J.** (2012). Toward the single-hour high-quality genome. *Annual Review of Biochemistry*, 81: 359-378.

Stewart, P. S. and Costerton, J. W. (2001). Antibiotic resistance of bacteria in biofilms. *Lancet,* 358: 135-138.

Stewart P.S., Hamilton M.A., Goldstein B.R. and **Schneider B.T.** (1996). Modeling biocide action against biofilms. *Biotechnology and Bioengineering*, 49: 445-455.

Stoodley P., Sauer K., Davies D.G. and Costerton J.W. (2002). Biofilms as complex differentiated communities. *Annual Review of Microbiology*, 56: 187-209.

Sutherland I.W. (2001). The biofilm matrix – an immobilized but dynamic microbial environment. *Trends in Microbiology,* 9: 222-227.

Su Y., Li B. and **Zhu WY** (2013). Fecal microbiota of piglets prefer utilizing DL-lactate mixture as compared to D-lactate and L-lactate in vitro. *Anaerobe,* 19: 27-33.

Szewzyk U., Szewzyk R., Manz W. and Schleifer K. H. (2000). Microbiological safety of drinking water. *Annual Reviews in Microbiology*, 54: 81–127.

Tachikawa M., Tezuka M., Morita M., Isogai K. and **Okada S.** (2005). Evaluation of some halogen biocides using a microbial biofilm system. *Water Research*, 39: 4126-4132.

Teske, A., Brinkhoff, T., Muyzer, G., Moser, D. P., Rethmeier, J., and **Jannasch, H. W.** (2000). Diversity of thiosulfate-oxidizing bacteria from marine sediments and hydrothermal vents. *Applied Environmental Microbiol*ogy, 66, 3125–3133

Thomas R.K., Nickerson E., Simons J.F., Janne P.A., Tengs T., Yuza Y., Garraway L.A., LaFramboise T., Lee J.C. (2006). Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Medicine*. ;12:852 – 855.

Tokajian S.T.H., Fuad A., Hancock I.C. and **Zalloua P.A.** (2005). Phylogenetic assessment of heterotrophic bacteria from a water distribution system using 16S rDNA sequencing. *Canadian Journal of Microbiology*, 51:325-335.

Tomaras A.P., Dorsey C.W., Edelmann R.E. and Actis L.A. (2003). Attachment to and biofilm formation on abiotic surfaces by acinetobacter baumannii: Involvement of a novel chaperone-usher pili assembly system. *Microbiology,* 149: 3473–3484.

Trudel L., St-Amand L., Bareil M., Cardinal P. and Lavoie M.C. (1986). Bacteriology of the oral cavity of BALB/c mice. *Canadian Journal of Microbiology,* 32: 673–678.

Tsai Y.P. (2006). Interaction of chlorine concentration and shear stress on chlorine consumption, biofilm growth rate and particle number. *Bioresource Technology*, 97: 1912-1919.

Tsai Y.P. (2005). Impact of flow velocity on the dynamic behaviour of biofilm bacteria. *Biofouling,* 21: 267-277.

USEPA (1992). Control of biofilm growth in drinking water distribution systems. Development, O.o.R.a. (ed), Washington.

Vaerewijck M.J., Huys G., Palomino J.C., Swings J. and Portaels F. (2005). Mycobacteria in drinking water distribution systems: Ecology and significance for

human health. *Federation of European Microbiological Societies Microbiology Reviews,* 29: 911–934.

Van der Kooij D., Veenendaal H.R. and Scheffer W.J.H. (2005). Biofilm formation and multiplication of Legionella in a model warm water system with pipes of copper, stainless steel and cross-linked polyethylene. *Water Research*, 39: 2789-2798.

Van der Kooij D. (2000). The unified biofilm approach: a framework for addressing biological phenomena in distribution systems. International Distribution Research Symposium. Denver, CO.

Van der Kooij D., Vrouwenvelder H.S. and **Veenendaal H.R.** (1995). Kinetic aspects of biofilm formation on surfaces exposed to drinking water. *Water Science and Technology.* 32: 61-65.

Van der Kooij D. (1992). Assimilable organic carbon as an indicator of bacterial regrowth. *Journal of American Water Works Association.* 84: 57–65.

Van Loosdrecht M.C.M., Picioreanu C. and **Heijnen J.J.** (1997). A more unifying hypothesis for biofilm structures. *Federation of European Microbiological Societies Microbiol Ecology.* 24: 181-183.

Van Loosdrecht M.C.M., Eikelboom D., Gjaltema A., Mulder A., Tijhuis L. and **Heijnen J.J.** (1995). Biofilm structures. *Water Science and Technology,* 32: 35-43.

Vaz-Moreira I., Nunes O.C. and **Manaia C.M.** (2012). Diversity and antibiotic resistance in *Pseudomonas* spp. from drinking water. *Science of the Total Environment,* 426: 366–374.

Vaz-Moreira I., Nunes O.C. and Manaia C.M. (2011). Diversity and antibiotic resistance patterns of *Sphingomonadaceae* isolated from drinking water. *Applied and Environmental Microbiology,* 77: 5697–5706.

Veiga M.C., Jain M.K., Wu W.M., Hollingsworth R.I. and **Zeikus J.G.** (1997). Composition and role of extracellular polymers in methanogenic granules. *Applied and Environmental Microbiology,* 63: 403-407.

Verberk J.Q.J.C., O'Halloran K.J., Hamilton L.A., Vreeburg J.H.G. and **Van Dijk J.C.** (2007). Measuring particles in drinking water transportation systems with particle counters. *Journal Water Supply Resources and Technology - AQUA* 56: 345-355.

Videla H.A. and **Characklis W.A.** (1992). Biofouling and microbially influenced corrosion. *International Biodeterioration and Biodegradation*, 29: 195–212.

Vieira M.J., Melo L. and **Pinheiro M.M.** (1993). Biofilm formation: hydrodynamic effects on internal diffusion and structure. *Biofouling,* 7: 67-80.

Volk C.J. and **LeChevallier M.W.** (2000). Assessing biodegradable organic matter. *Journal of American Water Works Association*, 92(5): 64-76.

Vornhagen J., Stevens M., McCormick D.W., Dowd S.E., Eisenberg J.N., Boles B.R., (2013). Coaggregation occurs amongst bacteria within and between biofilms in domestic showerheads. *Biofouling*, 29:53-68.

Vreeburg J. and **Boxall J.** (2007). Discolouration in potable water distribution systems: A review: *Water Research*, 41: 519–529.

Wadowsky R.M., Yee, R.B., Mezmar, l., Wing, E.J. and Dowling, J.N. (1982). Hot water systems as sources of *Legionella pneumophila* in hospital and nonhospital plumbing fixtures. *Applied and Environmental Microbiology,* 43:104-1 110

Wang Y.Y., Claeys L., van der Ha D., Verstraete W. and **Boon N.** (2010). Effects of chemically and electrochemically dosed chlorine on *Escherichia coli* and *Legionella beliardensis* assessed by flow cytometry. *Applied Microbiology and Biotechnology*, 87: 331–341

Wang Q., Garrity G.M., Tiedje J.M. and **Cole J.R.** (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16): 5261–5267.

Wasche S., Horn H. and **Hempel D.C.** (2002). Influence of growth conditions on biofilm development and mass transfer at the bulk/biofilm interface. *Water Research,* 36: 4775-4784.

Watnick P. and Kolter R. (1999). Steps in the development of a *Vibrio cholerae* El Tor biofilm. *Molecular Microbiology*, 34: 586-95.

Weinberger M., Berg S.L., Feuerstein I.M., Pizzo P.A. and Witebsky F.G. (1992). Disseminated infection with *Mycobacterium gordonae*: Report of a case and critical review of the literature. *Clinical infectious diseases,* 14: 1229–1239.

Wentzel M., Lötter L., Loewenthal R. and **Marais G.** (1986). Metabolic behaviour of *Acinetobacter spp.* in enhanced biological phosphorus removal- a biochemical model. *Water SA* 12: 209–224.

Whittake, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3): 280–338.

Whittaker R.H. (1972). Evolution and measurement of species diversity. *Taxon*, 21(2/3): 213–251.

Wijeyekoon S., Mino T., Satoh H. and **Matsuo T.** (2004). Effects of substrateloading rate on biofilm structure. *Water Research*, 38: 2479-2488.

Williams M.M., Santo Domingo J.W., Meckes M.C., Kelty C.A. and **Rochon H.S.** (2004). Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *Journal of Applied Microbiology,* 96: 954-964.

Wimpenny J. and **Colasanti R.** (1997). A unifying hypothesis for the structure of microbial biofilms based on cellular automation models. *Federation of European Microbiological Societies Microbiol Ecology,* 22: 1-6.

Wilczak A., Jacangelo J.G., Marcinko J.P., Odell L.H. and **Kirmeyer G.J.** (1996). Occurrence of nitrification in chloraminated distribution systems. *Journal-American Water Works Association,* 88: 74–85.

Williams M.M. and **Braun-Howland E.B.** (2003). Growth of *Escherichia coli* in model distribution system biofilm exposed to hypochlorous acid or monochloramine. *Applied and Environmental Microbiology,* 69: 5463–5471

Wilson M.V. and **Shmida A.** (1984). Measuring Beta Diversity with Presence-Absence Data. *Journal of Ecology,* 72: 1055-1064.

Wolfaardt G.M., Lawrence J.R., Robarts R.D. and **Caldwell D.E.** (1998). In situ characterization of biofilm exopolymers involved in the accumulation of chlorinated organics. *Microbial Ecology*, 35: 213-223.

World Health Organization (2009). Burden of disease and cost-effectiveness estimates. Water Sanitation and Health.

World Health Organization (2004). Geneva, Guidelines for Drinking Water Quality: 3rd Edition. Chapter 11: Microbial Fact Sheets.

Wright G.D. (2007).The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology,* 5, pp. 175–186

Wuertz S., Okabe S. and **Hausner M.** (2004). Microbial communities and their interactions in biofilm systems: an overview. *Water Science and Technology,* 49: 327-336.

Xi C., Zhang Y., Marrs C.F., Ye W., Simon C., Foxman B. (2009). Prevalence of antibiotic resistance in drinking water treatment and distribution systems. *Applied Environmental Microbiology,* 75 5714–5718

Xu J. (2011). Microbial Ecology in the Age of Metagenomics. In: Handbook of Molecular Microbial Ecology I. John Wiley & Sons, Inc., pp 111-122.

Xue Z., Hessler C. M., Panmanee W., Hassett D.J. and Seo Y. (2010). *FEMS Microbiol. Ecol.*, 2013, 83, 101–111Ceri H, Olson ME, Turner RJ. Needed, new paradigms in antibiotic development. *Expert Opinion in Pharmacotherapy*, 11:1233–1237.

Zanetti F., De Luca, G. and **Stampi, S.** (2000). Recovery of Burkholderia pseudomallei and B.
cepacia from drinking water. *International Journal Food Microbiol*ogy, 59**:** 67-72

Zhang M., Liu W., Nie X., Li C., Gu J., Zhang C. (2012). Molecular analysis of bacterial communities in biofilms of a drinking water clearwell. *Microbes Environment,* 27: 443-448.

Zhang P., LaPara T.M., Goslan E.H., Xie Y. and **Parsons S.A.** (2009). Biodegradation of haloacetic acids by bacterial isolates and enrichment cultures from drinking water systems. *Environmental Science and Technology,* 43: 3169–3175.

Zhang T.C. and **Bishop P.L.** (1994). Structure, activity and composition of Biofilms. *Water Science and Technology,* 29: 335-344.

APPENDICES

Appendix I: Map of Kenyan Coastal Region

Appendix II: List of primers used for 16 S rRNA gene amplification

Sample ID	Barcode Sequence	Linker Primer Sequence
BRB.1	ACTCACAG	AGRGTTTGATCMTGGCTCAG
BRB.2	ACTCACTC	AGRGTTTGATCMTGGCTCAG
BRB.3	ACTCCAGA	AGRGTTTGATCMTGGCTCAG
BRB.4	ACTCGAGT	AGRGTTTGATCMTGGCTCAG
BRW.1	ACGTTGGT	AGRGTTTGATCMTGGCTCAG
BRW.3	ACTCCACT	AGRGTTTGATCMTGGCTCAG
BRW.4	ACTCGACA	AGRGTTTGATCMTGGCTCAG
BRW.5	ACTCTCAC	AGRGTTTGATCMTGGCTCAG
BRW.6	ACTCTCTG	AGRGTTTGATCMTGGCTCAG
MZB.1	ACGTGATC	AGRGTTTGATCMTGGCTCAG
MZB.2	ACGTGTTG	AGRGTTTGATCMTGGCTCAG
MZB.4	ACGTTGCA	AGRGTTTGATCMTGGCTCAG
MZB.5	ACTCAGTG	AGRGTTTGATCMTGGCTCAG
MZB.7	ACTCGTGA	AGRGTTTGATCMTGGCTCAG
MZW.1	ACGTGAAG	AGRGTTTGATCMTGGCTCAG
MZW.2	ACGTGTAC	AGRGTTTGATCMTGGCTCAG
MZW.3	ACGTTCCT	AGRGTTTGATCMTGGCTCAG
MZW.4	ACGTTCGA	AGRGTTTGATCMTGGCTCAG
MZW.5	ACTCAGAC	AGRGTTTGATCMTGGCTCAG
MZW.6	ACTCCTGT	AGRGTTTGATCMTGGCTCAG
MZW.7	ACTCGTCT	AGRGTTTGATCMTGGCTCAG

Appendix III: The proportion of phyla phylotypes recovered in distribution systems samples

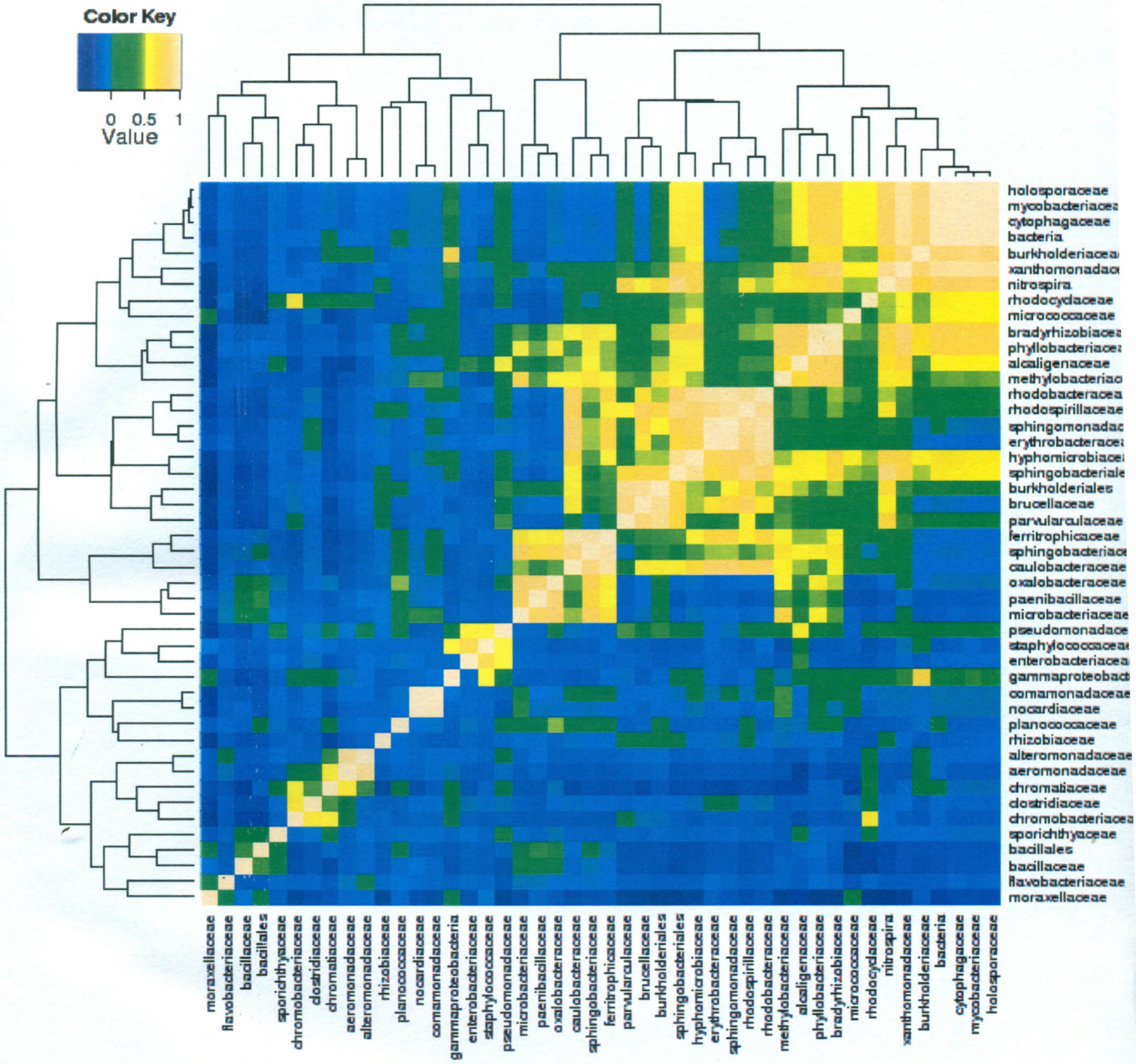
Taxon	BRW.1	BRB.1	MZB.1	MZW.2	BRW.5	BRB.2	MZW.7	MZW.4	MZW.5	BRW.6	MZB.7
Acidobacteria	0	0.0033	0.007194	0	0.014184	0.005	0.005587	0	0	0	0
Actinobacteria	0	0.013201	0.172662	0	0.06383	0	0	0	0	0	0.015504
Armatimonadetes	0	0	0	0	0.004728	0	0	0	0	0	0
Bacteria;BD1-5	0	0	0	0	0.004728	0	0	0	0	0	0
Bacteroidetes	0.054545	0.059406	0.014388	0	0.12766	0.03	0.022346	0.039216	0	0	0.023256
CD BRC1	0	0	0	0	0.002364	0	0	0	0	0	0
CD OD1	0	0.046205	0	0	0.01182	0.005	0	0	0	0	0
CD OP11	0	0.009901	0	0	0.002364	0	0	0	0	0	0
CD OP3	0	0	0	0	0.004728	0	0	0	0	0	0
CD TM7	0	0.072607	0	0	0.004728	0	0	0	0	0	0
CD WS3	0	0	0	0	0.002364	0	0	0	0	0	0
Chlorobi	0	0.0033	0	0	0	0	0	0	0	0	0
Chloroflexi	0	0.0033	0.028777	0	0.037825	0	0	0	0	0	0.007752
Cyanobacteria	0	0.0033	0.007194	0	0.002364	0.02	0	0	0	0	0
Deinococcus-Thermus	0	0	0	0	0	0	0	0	0	0	0
Elusimicrobia	0	0	0	0	0	0.045	0	0	0	0	0
Firmicutes	0.127273	0.0033	0.201439	0.337349	0.002364	0.005	0.089385	0.27451	0.5	0	0.224806
Gemmatimonadetes	0	0.0033	0.014388	0	0.004728	0	0	0	0	0	0
Lentisphaerae	0	0	0	0	0	0	0	0	0	0	0
NPL-UPA2	0	0.006601	0	0	0.007092	0.005	0	0	0	0	0
Nitrospirae	0	0.013201	0.014388	0	0.009456	0.02	0	0	0	0	0

Planctomycetes	0	0.009901	0.007194	0	0.130024	0.015	0	0	0	0	0.007752
Proteobacteria	0.818182	0.729373	0.52518	0.662651	0.50591	0.845	0.882682	0.686275	0.5	1	0.72093
SM2F11	0	0.006601	0	0	0	0	0	0	0	0	0
TM6	0	0	0	0	0	0.005	0	0	0	0	0
Verrucomicrobia	0	0.013201	0	0	0.056738	0	0	0	0	0	0
Other	0	0	0.007194	0	0	0	0	0	0	0	0

Taxon	**MZW.1**	**MZB.2**	**BRW.4**	**BRB.3**	**MZW.3**	**MZW.6**	**MZB.5**	**MZB.4**	**BRB.4**
Acidobacteria	0	0	0	0	0	0	0.016129	0.043478	0.014925
Actinobacteria	0.057143	0	0	0.014286	0	0	0.032258	0.065217	0.014925
Armatimonadetes	0	0	0	0	0	0	0	0	0
Bacteria;BD1-5	0	0.01	0	0	0	0	0	0	0
Bacteroidetes	0.128571	0.07	0.125	0.028571	0.051724	0	0	0.369565	0.059701
CD BRC1	0	0	0	0	0	0	0	0	0
CD OD1	0	0	0	0	0	0	0	0	0
CD OP11	0	0	0	0	0	0	0	0	0
CD OP3	0	0	0	0	0	0	0	0	0
CD TM7	0	0	0	0	0	0	0	0	0
CD WS3	0	0	0	0	0	0	0	0	0
Chlorobi	0	0	0	0	0	0	0	0	0
Chloroflexi	0	0	0	0	0	0	0	0	0
Cyanobacteria	0	0	0	0	0	0	0	0.021739	0
Deinococcus-Thermus	0	0	0	0	0	0	0	0	0
Elusimicrobia	0	0	0	0	0	0	0	0	0

Firmicutes	0.128571	0.05	0.048077	0.171429	0.086207	0.4375	0.016129	0	0.074627
Gemmatimonadetes	0	0	0	0	0	0	0	0	0
Lentisphaerae	0	0	0	0	0	0	0	0	0
NPL-UPA2	0	0	0	0	0	0	0	0	0
Nitrospirae	0	0	0	0	0	0	0	0	0
Planctomycetes	0	0	0	0	0	0	0.016129	0	0.014925
Proteobacteria	0.685714	0.87	0.817308	0.785714	0.862069	0.5625	0.919355	0.478261	0.820896
SM2F11	0	0	0	0	0	0	0	0	0
TM6	0	0	0	0	0	0	0	0	0
Verrucomicrobia	0	0	0	0	0	0	0	0.021739	0
Other	0	0	0.009615	0	0	0	0	0	0

Appendix IV: Heat map of the OTUs occurring in all Mzima and Baricho water lines



Each column represents a different sample group; each row represents a different OTU; the color of the cells represents the frequency of that OTU within the sample group (percentage of the group samples the OTU was retrieved from).

The phylogenetic data indicated that the microbial community is constructed out of a significant number of different and mostly nonpathogenic proteobacterial species. Of these many are probably novel and have not yet been cultured.