A METAGENOMIC INSIGHT INTO THE ROLE OF WASTEWATER TREATMENT PLANTS AS POTENTIAL HOTSPOT FOR ANTIBIOTIC RESISTANT BACTERIA AND ANTIBIOTIC RESISTANCE GENES

by

PO-CHENG TANG

Submitted in fulfilment of the academic requirements for the degree of Master of Science (MSc) in the Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science at the University of KwaZulu-Natal, Durban, South Africa.

June 2016

As the candidate's supervisor, I have approved this dissertation for submission.

Name: Prof. A. O. Olaniran Signed: _____ Date: _____

Abstract

The magnitude of the global freshwater crisis is underestimated. Although international and national efforts have implemented strategies to safeguard these precious resources, the consequential effect of continued deterioration of water quality on the available options for water usage is still a major area of concern. Contamination of surface water resources resulting from inadequate treated effluent discharge from wastewater treatment plants (WWTPs) has previously been indicated as an important topic requiring continued investigations. Often, these discharges introduce large amounts of organic matter and nutrients which could lead to eutrophication and temporary oxygen deficiencies, ultimately disrupting the natural biotic community structure and its important ecological functions. In addition, pathogenic bacteria and antibiotic resistance genes (ARGs) present in the wastewater discharged increases public health risks due to their disease causing potential and dissemination of resistance determinants, respectively. This could result in increased cases of diseases such as dysentery, cholera, skin infections and typhoid, leading to higher rates of mortality and morbidity, placing further stresses on the public health system. Hence, better management and minimisation of microbial pathogens and antibiotic resistant bacteria in WWTPs is crucial to prevent the dissemination of potential pathogens and ARGs into the environment. The overall objective of this study was to provide metagenomic insights into the bacterial diversity as well as the ARGs associated with the communities present in wastewaters and to establish the impact of the discharged effluents on the receiving river bodies. Samples were collected from two urban WWTPs receiving distinctive raw sewage, with combinations of domestic, industrial and hospital, in the city of Durban, South Africa. Metagenomic sequencing were conducted on the Roche 454 platform and shotgun generated pyrosequences were analysed using established pipelines

(MG-RAST) and a plethora of bioinformatic tools (IDBA_UD, MaxBin, CheckM, etc.) and databases (VFDB and CARD). The findings of this study suggest that there is a substantial shift between the bacterial communities in the raw wastewater and the treated wastewater in terms of abundance, composition and diversity. Similarly, this was observed in the functional genes determined and the overall metabolic potential of the associated bacterial communities. Additionally, putative genes encoding for resistance to most classes of antibiotics were identified in all samples and encompassing the three major resistance mechanisms. Although shotgun sequencing and subsequent analysis only provides qualitative and semi-quantitative data, we were able to establish a list of the bacterial communities and antibiotic resistance genes across all the water samples, identified the potentially pathogenic bacteria causing human disease in the treated wastewaters of both plants that need to be considered during treatment management decisions, as well as the functional potential of the communities. Overall, results from this study demonstrated the usefulness of a metagenomic approach in initial stages of water quality assessments. The results of this work also provide significant and novel information that will contribute to our understandings of water quality which could lead to sustainable wastewater management decisions.

"If you take care of the small things, the big things take care of themselves."

– Emily Dickinson (1830-1886)

Preface

The experimental work described in this dissertation was carried out in the Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal (Westville campus), Durban, South Africa from February 2014 to August 2015, under the supervision of Professor Ademola O. Olaniran.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

Declaration 1 - Plagiarism

I, Po-Cheng Tang declare that:

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written but the general information attributed to them has been referenced
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: _____

Declaration 2 - Publications and Conferences

For all publications, P. Tang and A.O. Olaniran both conceptualized and designed the research project as well as contributed equally in the preparation of all the listed manuscripts; P. Tang implemented and analyzed experimental work.

Publication 1: P. Tang and A.O. Olaniran (2016). Fate of bacterial pathogens, community composition and antibiotic resistance genes in two full-scale wastewater treatment plants as revealed by metagenomic analysis. *Science of the Total Environment* (Submitted).

Publication 2: P. Tang and A.O. Olaniran (2016). Chlorination effects on bacterial diversity, functional potential and antibiotic resistome during urban wastewater treatment: A metagenomic insight. *Water Research* (Submitted).

Publication 3: P. Tang and A.O. Olaniran (2016). Antibiotic resistance genes, bacterial community structure and metabolic potential in two rivers impacted by wastewater treatment plant discharges: a metagenomic approach. *Environmental Science and Pollution Research* (Submitted).

Publication 4: P. Tang and A.O. Olaniran (2016). Bacterial community dynamics of two rivers receiving treated urban wastewater effluents: a T-RFLP assessment. *World Journal of Microbiology and Biotechnology* (Submitted).

Publication 5: P. Tang and A.O. Olaniran (2016).Draft genome sequence of an *Acinetobacter* sp. PT1 reconstructed from metagenomic sequences of wastewater effluent samples from a wastewater treatment plant in Durban, South Africa. *Genome Announcement* (in preparation).

Publication 6: P. Tang and A.O. Olaniran (2016). Draft genome sequence of a *Polynucleobacter necessarius* PT2 reconstructed from metagenomic sequences of wastewater effluent samples from a wastewater treatment plant in Durban, South Africa. *Genome Announcement* (in preparation).

Declaration 2 - Publications and Conferences

Conference Presentations: P. Tang and A.O. Olaniran (2015). Chlorination effects on bacterial diversity and antibiotic resistance gene profiles of treated wastewater: A metagenomic insight. *VI International Conference on Environmental, Industrial and Applied Microbiology (BioMicroWorld 2015); Barcelona, Spain.*

Signed: _____

Contents

List of Figures

List of Tables

Acknowledgements

More than two years later, here I am ready for the next chapter in my scientific adventures. There is quite a lot of people whom one way or the other played a role my journey thus far and I am grateful to all of them!

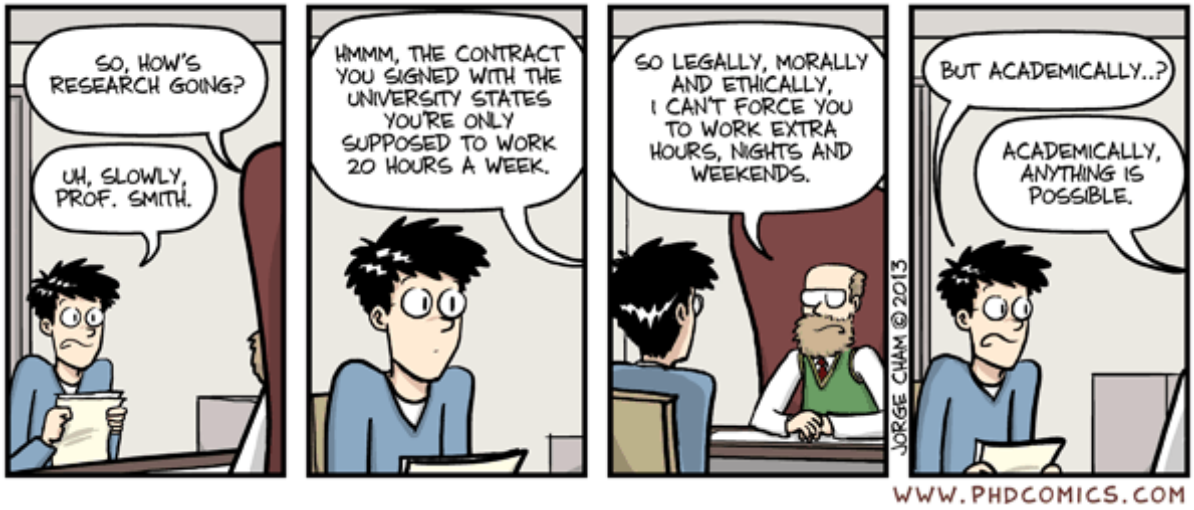
First of all, I would like to thank my supervisor **Prof. Ademola 'Ade' Olaniran** for believing in me enough to give me the opportunity to pursue this degree in his research lab. Prof., there's so much to say but I honestly am not sure where to begin. So allow me to summarize with a quote from Isaac Newton, *"If I have seen further than others, it is by standing upon the shoulder of giants"*. I learned a great deal from you during my graduate studies, and the most important thing was the way you've always managed to guide me to see the bigger picture and to put things into perspective. This is something that I wish to carry with me in my upcoming adventures through life and I'll try my best to also pass this onto others.

I am grateful to my friends, colleagues, 'co-supervisors', and just bloody awesome Ashmita Arjoon, Lekita Singh and Joash 'Govindesamy'. Ash, Lex and Joey, we have come a long way since what seems like a decade, although it has only been just over three years since the four of us met in the disturbingly grey Laboratory 4. **Ash**, from the ridiculous early mornings where we were able to see the sun rise over Reservoir Hills at the parking lot for undergrad tutorials, to the extremely late nights' shenanigans, running and staining agarose gels, pumping the volume on club songs to keep us sane from filtering those waters (just thinking about it is giving me a headache, though, at least we had the night-view from the windows). Of course, to have traveled and explored literally all over Europe and marveled at what the world has to offer our seeming trifling lives, with you and your endless 'selfies', has been such a wonderful time it saddens me that this period seems to be coming to an end. Lex and Joey, both of your willingness to assist me in the project is something that I will never forget and I'm truly thankful for your help in 'saving the day'! **Lex**, it's been more than six years since I've known you and I don't think this entire journey would have been the same without you. We've really done a marathon here and I'm excited to see where things go. I really enjoyed the fact that we were able to talk about anything and everything with no limits, from the guys at the department

thinking we were a couple, given the rules on dating colleagues in a working environment, to freezing 'dem' eggs, it cracks me up just thinking about it. **Joey**, where do I even begin my brother from another mother. I'm glad that we got along really well and I'm always grateful for your 'serious yet not so serious' chats and advice on life. You definitely kept the lab together, although I had always questioned why you did so, but now I can see why you did so in retrospect. I wish you all the best on all your entrepreneurial endeavors and just remember us when you've gotten your own private jet and island. Though this paragraph would never be able to express how thankful I am to you three marvelous creatures of the world for keeping me grounded and providing me with liters of much needed endorphins, I hope we may never forget the past few years.

To Megan, Lucrecia, Tom, Ashley, Renee and the past and current members of the **laboratory 2** research group, it was always such a pleasure to escape to the 'safe heaven'. **Meg**, it's ridiculous how much has transpired since the time we've gotten to know each other and I honestly appreciate your straightforwardness and remarks about anything and everything. I wish you all the best on your new scientific adventure at DKFZ and know that you'll be great there. **Lu**, your smile always made my day and I really appreciate the coffee and tea sessions with you, it was such a pleasure and I'm sure the long walk to freedom is near so hang in there. Just think of the exciting future that's installed. **Tom**, **Ash, Renee** and **Chika**, allow me to use this meme to summarize:

To the current and previous members of the **laboratory 4** research group, I am appreciative of you guys for putting up with my craziness all these years. It was such a pleasure to have gotten to know some of you and though I'm sure at times research must not have been easy, just remember caffeine and friends definitely helps, but only to a certain extent before you've completely been 'zombified'. Best of luck to everyone in your research and hopefully this comic will help with the much needed endorphins:



Thanks to all the **academic, technical and administrative staff** at the Discipline of Microbiology at the University of KwaZulu-Natal (Westville Campus) for your support, assistance and ensuring everything works as smoothly as possible throughout the duration of the research project.

I would also like to thank the South African **National Research Foundation** (NRF) and **Medical Research Council** (MRC) for their financial support during this research project. Furthermore, thanks to the NRF Travel Grant and **Professor Daniel Huson** for hoisting my brief visit to the University of Tübingen. Learning from you, **Sina**, **Anna**, **Rewati** and **Mohomed** on a what seem to be a '180-degree' field from microbiology at the time was such a wonderful experience. You guys definitely helped in guiding me towards loving this new interdisciplinary field of genomics and I am sure we're bound to bump into each other in the future. To the Erasmus Mundus INSPIRE scholarship and **Professor Dan Andersson** for accommodating me during my visit to Uppsala University and agreeing to provide me with the opportunity to learn from your research group even though it was a special case. **Erik L.**, **Erik G.**, **Marius**, **Pikkei, Jessica, Lisa** and the DA and DH groups, thank you for making my time at Uppsala such a wonderful experience, from the card games to champagne celebrations at the fika corner (thanks for getting the non-alcoholic version for me), it was always a pleasure to have lunch with you guys and chat about anything to do with daily life as a researcher. Learning from you guys was such a rewarding and such eye opening experience which not only enriched my personal growth but also showed me the different scientific adventures one could take.

Michelle, **Karen**, **Tim**, **Tegan**, **Mpume**, **Chantal**, **Henrik** and **Akoth** I am thankful for your friendship and the wonderful memories that we have together in Sweden. Your support in the

last leg of this research project will forever be appreciated and I'm looking forward to what the future has arranged for all of us!.

Lastly, to my parents, **Derek** and **Janice**, I have and always will be grateful for your never ending support when it comes to my tertiary studies. This degree and any achievements thus far would not have been possible without it and is especially dedicated to you both. My brother **James**, for silently being there for me during this long period of my studies. I am forever indebted to your kindness.

Chapter One

1 Introduction and Literature Review

1.1 Introduction

Water is a natural resource and the driving force of nature by virtue of its fundamental properties. Clean and safe drinking water is indispensable for sustenance, health and dignity of life, whilst freshwater is central for energy and food security as well as all ecosystems functions. Hence, our societies and continued development profoundly depend on this natural resource. Yet, the magnitude of the global freshwater crisis is underestimated. This is partly due to the ignorance of the fact that unlike other natural resources, freshwaters cannot be regarded as a constant supply, but should rather be regarded as a restricted flux, which needs to be correctly managed so that it continues to remain as a renewable resource (World Economic Forum, 2015).

To this end, the United Nations (UN) General Assembly in December 2003, in resolution A/RES/58/217, proclaimed the period of 2005 to 2015 the International Decade for Water of Life Action and its' Environment Programme (UNEP) has acknowledged that water-related problems are one of the most immediate and serious environmental threats to humankind (UNEP, 2005). During this period, the importance of this finite resource was stressed at the Bonn Conference in 2011, and again in 2012 at The Rio+20 Summit in conjunction with global energy and food security issues in sustaining a Green Economy (Jägerskog and Jønch Clausen, 2012). Moreover, the Millennium Development Goals (MDG), which came to a conclusion at

the end of the 2015, were implemented by most countries across the globe with success of the target for safe drinking water met early in 2010 (UN, 2015a).

Despite meeting the goal seven's target in providing two billion people access to safe drinking water sources, and the official launch of the bold and transformative 2030 Agenda for Sustainable Development this year, with a new set of 17 Sustainable Development Goals (SDG) established for the next 15 years, depletion of existing water resources still continues to be a major problem, with water scarcity affecting more than 40% of the global population and has been projected to increase with the explosive population growth (UN, 2015b). This is in part due to the limitation of the MDG framework in addressing the full water and development agenda as well as contributing factors such as increasing poverty, rapid urbanization, complex hydro-geological variability and the effects of climate change (Haddeland *et al.*, 2014). Additionally, the crisis is further compounded by constant national, international and trans-boundary conflicts arising in an attempt to provide adequate food, water and health security for the entire populations (Kraljevic and Geiger, 2012). Moreover, lack of potable water has been estimated to cost countries between 1% to 7% of their annual Gross Domestic Product, impeding national economic growth (Fogden and Wood, 2009). Furthermore, environmental considerations and its synergistic role with socio-economical dynamics were scarcely addressed. Combined, these aspects place further stress on the already deteriorating water and sanitation infrastructure, more so in mid-lateral developing countries than developed countries, where billions of people are already at risk of water, sanitation and hygiene (WSH) related diseases (WHO/UNICEF, 2012).

Fundamentally, the available abundance of freshwater sources, occurring as surface waters (ice sheets, ice caps, glaciers, icebergs, bogs, ponds, lakes, rivers and streams) and ground waters (underground aquifers and streams), has been suggested to be enough for sustaining human societies (Bunn, 2016). Approximately 93 000 km^3 is estimated to be readily available from rivers, streams, ponds and lakes with more than double the amount suggested to be locked as glaciers and ice or stored in ground waters (Bunn, 2016). Stockholm's International Water Institute estimated that, on average, each person on Earth requires a minimum of 1 000 m^3 of water per year for drinking, hygiene and growing food for sustenance, which is equivalent to two-fifths of the volume of an Olympic-size swimming pool (Rockström *et al.*, 2007). However, these sources are not uniformly distributed in time and space, nor in relation to anthropological settlement and urbanization patterns (Bunn, 2016). In practice, ground water contributes a large proportion of freshwater used across the domestic, agricultural and industrial sectors. Here, agriculture is the greatest consumer, accounting for approximately 70% of the total consumption, whilst industries consume 25% and the remaining 5% is used for domestic and recreational purposes. Though, it should be noted that these estimates are dependent on the country in question and their development stage, with as high as 95% total consumption in agriculture alone in developing countries. This high and projected increase in percentage consumption from the agricultural sector is a resultant chain effect of the explosive population growth leading to increases per capita of food consumption. Hence, the enormous expansion of the agriculture sector and the indirect role of ground water in sustaining life contributes to their large demand (UNEP, 2008). This has led to the depletion of major ground water resources, resulting from its overuse, improper management as well as pollution from point and non-point sources. As a consequence, nearly 80% of the world's population has been estimated to be exposed to high levels of threat to water security (Vörösmarty *et al.*, 2010). The above factors serve as the major driving force behind the increased use of wastewater, rain

water, grey water, storm water and surrounding surface waters as an alternative resource for the agricultural, domestic and industrial sectors (WHO/IWA, 2011).

Although, the importance of these alternative water sources has become clear over the past decade, pollution and contamination of these sources from point (municipal and industrial) and/or non-point (agricultural, mining, sewage, domestic) sources has become increasingly evident over the past few decades. Water polluting substances including organic chemicals (e.g. phenolic compounds), inorganic chemicals (e.g. phosphates) and heavy metals (e.g. mercury), are frequently used in both the industrial and agricultural sectors (Beesley *et al.*, 2011; Guo *et al.*, 2012; Stefanowicz *et al.*, 2009), their contamination of water and subsequent release into the environment results largely from the non-compliance of mines, industrial factories, wastewater treatment plants and crop cultivation productions, to the effluent quality regulations, good agricultural practices and disposal guidelines stipulated by government (Olaniran and Igbinosa, 2011; FAO, 2004; Ayomoh *et al.*, 2008). Several studies have suggested the agricultural sector as one of the major contributors to water pollution (Liu *et al.*, 2013; Ribbe *et al.*, 2008) owing to the large use of synthetic antibiotics, fungicides, pesticides, insecticides and herbicides to ensure maximum production with low costs (Kughur, 2012). Pollution of nearby surface and ground water by these chemicals has been shown to result from the effects of leaching, runoff and infiltration (Dolan *et al.*, 2012; Mendes *et al.*, 2012; Masrevaniah, 2010; Shanafield *et al.*, 2010). Mortality attributed to WSH related diseases is a global concern, with approximately 50% of the reported cases arising from microbial intestinal infections (Cabral, 2010). According to the World Health Organization (WHO), such diseases account for an estimated 4.1% of the total disability-adjusted life year measurement of the global burden of disease and approximately 3.4 million people have been estimated, mostly children, to die from

water-related diseases each year, whilst the UN Children's Fund (UNICEF) assessment reported death cases of 4 000 children each day as a result of contaminated water (WHO, 2014a). To tackle this problem, the WHO has repeatedly stressed the importance of improving water quality on the reduction of the global disease burden which lead to the establishment of the MDG seven targets (Pandey *et al.*, 2014). Naturally, microorganisms are ubiquitous and variety of these microbes including bacteria, algae, fungi, protozoa and viruses, exist in water sources where they form an ecosystem in which the dynamics are complex and difficult to comprehend. Though, all water sources can become contaminated from natural geogenic activities, most waterborne diseases result largely from some form of faecal pollution. Therefore, reliable water treatment infrastructures and continuous water monitoring are essential towards containment of anthropogenic activities and minimizing its influence on the environment, ensuring good water quality, community sanitation and health as well as lowering risks of epidemics and outbreaks (Cabral, 2010).

1.2 Microbial water quality assessments

For decades, to ensure safety of water sources such as those described above as well as those for drinking and recreational activities, water quality assessments have largely been based on *in vitro* cultivation, enumeration and detection of representative indicator microorganisms or the disease causing pathogens (Carew *et al.*, 2013). The occurrence and abundance of indicator organisms serve as proxies, as they are easily measured and could be correlated to agents that directly mediate waterborne risks. Use of indicator organisms over direct pathogens is due to the time consuming, costly and technical difficulties associated with the detection of all

possible pathogenic organisms present in water sources as well as the current understandings of the presence and absence of these representative indicator organisms relative to disease causing pathogens (Barrell *et al.*, 2000; Leclerc *et al.*, 2002). It is important to note that these indicator microbes are flawed as they could also be derived from non-human sources and may be subjected to ecological or environmental stresses which have been shown to compromise their suggestive power for its associated pathogens (Payment and Locas, 2011). Nonetheless, indicator organisms are often used to predict the origin or source of contamination, coined "microbial source tracking" (Shanks *et al.*, 2010). For example, a study demonstrated microbial source tracking using at least two species within the *Bacteroides* genus as indicators of human contamination and showed their direct association with symbionts of the human gut microflora (Yampara-Iquise *et al.*, 2008). In order to eliminate the ambiguity with the term "microbial indicators", the WHO (2001) has defined three possible groups described in **Table 1.2**. Furthermore, an ideal indicator organism should be suitable for all categories of water, be present with the occurrence of the associated pathogen, present in higher numbers than its associated pathogen, should have survival characteristics similar to its associated pathogen, should not be pathogenic, should not multiply in waters, and easily, rapid and cost-effectively quantifiable even at low numbers (Horan, 2003). As no single organism is able to fulfil all of the above mentioned characteristics, several indicator organisms have been selected (with a few key faecal indicators briefly described in **Table 1.3)**, each with certain characteristics and are still routinely used for water quality and health risk assessments with their presence serving as an indication of the occurrence of its associated pathogens. Alternatively, specific pathogens such as waterborne *Legionella*, *Campylobacter*, *Leptospira*, viruses and protozoa (more described in **Table 1.1**) are often directly detected and enumerated over the use of indicator organisms. Direct detection and enumeration allows for quantitative microbial risk assessment,

where environmental concentrations of pathogens are compared to models of infectivity in order to characterize the risk to exposed human populations (WHO, 2001).

Table 1.1: Microbial pathogens associated with waterborne diseases (adapted from WHO, 2011).

Pathogen	Diseases	Health significance	Persistence in water supplies	Resistance to chlorine	Relative infectivity	Important animal source
Bacteria						
Burkholderia pseudomallei	Melioidosis	High	May multiply	Low	Low	No
Campylobacter spp.	Gastroenteritis	High	Moderate	Low	Moderate	Yes
Escherichia coli (pathogenic & enterohaemorrhagic)	Gastroenteritis	High	Moderate	Low	Low to High	Yes
Francisella tularensis	Tularemia	High	Long	Moderate	High	Yes
Legionella spp.	Legionellosis	High	May multiply	Low	Moderate	No
Leptospira spp.	Leptospirosis	High	Long	Low	High	Yes
Mycobacteria spp. (non-tuberculous)	Chronic obstructive pulmonary disease	Low	May multiply	High	Low	No
Salmonella Typhi	Thyphoid fever	High	Moderate	Low	Low	No
Other *Salmonella* spp.	Salmonellosis/ gastroenteritis	High	May multiply	Low	Low	Yes
Shigella spp.	Dysentery	High	Short	Low	High	No
Vibrio cholerae	Cholera	High	Short to long	Low	Low	No
Viruses						
Adenoviruses	Gastroenteritis	Moderate	Long	Moderate	High	No
Astroviruses	Gastroenteritis	Moderate	Long	Moderate	High	No
Enteroviruses	Pleurodynia, Hand-foot-and-mouth disease, Herpangina, Poliomyelitis	High	Long	Moderate	High	No
Hepatitis A & E virus	Hepatitis	High	Long	Moderate	High	Potentially to No
Noroviruses	Gastroenteritis	High	Long	Moderate	High	Potentially
Rotaviruses	Diahorrea, vomiting	High	Long	Moderate	High	No
Sapoviruses	Gastroenteritis	High	Long	Moderate	High	Potentially
Protozoa						
Acanthamoeba spp.	Acanthamoeba keratitis, Granulomatous Amebic Encephalitis	High	May multiply	High	High	No
Cryptosporidium hominis/parvum	Diahorrea	High	Long	High	High	Yes
Cyclospora cayetanensis	Cyclosporiasis	High	Long	High	High	No
Entamoeba histolytica	Amebic dysentery	High	Moderate	High	High	No
Giardia intestinalis	Giardiasis	High	Moderate	High	High	Yes
Naegleria fowleri	Primary amebic meningoencephalitis	High	May multiply	Low	Moderate	No

Table 1.2: Definitions for indicator and index microorganisms of public health concern (adapted from WHO, 2001).

Group	Definition
General/process indicators	A group of organisms that demonstrates the efficacy of a process, such as total heterotrophic bacteria or total coliforms for chlorine disinfection.
Faecal indicators	A group of organisms that indicates the presence of faecal contamination, such as the bacterial groups thermo-tolerant coliforms or *E. coli*. Hence, they only infer that pathogens may be present.
Index organisms and model organisms	A group/or species indicative of pathogen presence and behaviour respectively, such as *E. coli* as an index for *Salmonella* and F-RNA coliphages as models of human enteric viruses.

Specifically, initial methods for identification of these indicators or pathogens rely on selective culturing and enumeration of presumptive isolates, followed by species confirmation using biochemical methods, such as the indole, methyl red, Voges-Proskauer (IMViC) tests. Such efforts lead to the standardization of simple and routine procedures for quantification of coliform indicators, which are relatively easy to culture (Simpson *et al.*, 2002). Conversely, cultivation methods for detection of specific pathogens are often difficult and time-consuming, with failure of detecting some organisms due to fastidious growth as in the case for *Leptospira* species (Evangelista and Coburn, 2010). In addition, some pathogens which are viable but non-culturable (VBNC) cannot be isolated by standard cultivation-based methods, meaning they are known to exist in a dormant infective state.

Hence, to overcome this limitation, cultivation-independent molecular methods have been developed and used in conjunction with conventional methods for the identification of indicator organisms and disease- causing pathogens. These methods target DNA extracted from water and other environmental samples which is subsequently subjected to analysis for the presence

and abundance of unique genes associated with the indicator species or pathogens of interest. This is entirely evident in terms of the frequently used small sub-unit of the ribosomal RNA gene (SSU rRNA), the 16S rRNA gene for Bacteria and 18S rRNA gene for Eukarya, for molecular-based analysis due to its ubiquity in all organisms as well as highly conserved variable and hyper-variable regions in the sequence structure (Green *et al.*, 2014). This has in turn allowed for alignment of multiple DNA sequence across diverse organisms and subsequent taxonomic identification and placement on the tree of life (Pace, 2009). As these cultivation-independent methods are either based on the direct detection of nucleic acids or PCR-based amplification of target genes, limitations associated with culture-dependent methods have been overcome, though not with its own set of challenges and limitations. Primarily, the quantification and detection of DNA extracted from environmental species has and is still often assumed to be derived from living organisms. However, free or naked DNA may also be isolated from dead organisms during methods in DNA extraction which doesn't take this into account. This ultimately hinders the correlation between cell abundance to the frequency of DNA copies isolated. Secondly, there is the limitation with regards to some methods in their ability to detect trace levels of targeted nucleic acids, such as virulence factors. Thirdly, quantification of microbial risks with the use of DNA-based methods may be complicated by the fact that pathogens have highly dynamic and diversified genomes, with additional complexities at the strain level, such as strain specific virulence factors (Tan *et al.*, 2015b). Therefore, quantification of pathogenic organisms based on the occurrence of a DNA biomarker, such as the use of the universal SSU rRNA gene, does not necessarily translate into public health risk should the strain detected lack virulence genes to cause diseases. Nonetheless, PCR-based methods have still been proven useful with some methods currently under evaluation (Green *et al.*, 2014) and one method already approved for the quantification of *Enterococcus* by the U.S. Environmental Protection Agency (USEPA, 2013, 2009). In general,

as cultivation-independent methods have now moved away from its infancy and is currently adopted in routine laboratory analysis, advancement in future water quality analysis hinders on the importance of defining a set of standard guidelines as well as thresholds of acceptable risk.

Table 1.3: Definitions of selected key faecal indicator micro-organisms (adapted from WHO, 2001).

Indicators	Description
Coliforms	Early water microbiologists defined all coliform bacteria as those bacteria which is able to grow within 24 to 48 hours at 36 ± 2 °C in the presence of bile salts, which is used to inhibit non-intestinal bacteria, and has the ability to produce acid and gas with the fermentation lactose. Faecal coliforms were considered to be those coliforms which were exclusively of faecal origin and consequently able to grow, replicate and ferment lactose at an elevated temperature of 44.5 ± 0.2 °C within 24 ± 2 hours, in addition to the production of indole from tryptophan metabolism (WHO, 2011). Hence, the coliform group includes species belonging to the genera *Citrobacteri*, *Enterobacter*, *Escherichia*, *Hafnia*, *Klebsiella*, *Serratia*, and *Yersinia*. Though some members of the *Aeromonas* genus may mimic coliforms in standard laboratory tests for the latter, they are oxidase-positive compared to oxidase-negative coliforms (Horan, 2003).
Escherichia coli	*E. coli* is commonly regarded as one of the first microbe in water monitoring programs and serves as the general primary indicator for faecal contamination of water. This species falls under the thermophilic category of coliforms which is able to produce indole from tryptophan. Their primary detection in water analysis is due to their prevalence in the gut microflora as well as large numbers excreted from both human and warm-blooded animals (WHO, 2001). Though majority of *E. coli* strains are harmless and are important components of healthy microflora, six major pathogenic classes has been established to date and are collectively referred to as diarrheagenic *E. coli* due to their resulting in diarrhoea, *viz.* enterotoxigenic *E. coli* (ETEC), enterohaemorrhagic, verocytotoxin-producing or shiga-toxin producing *E. coli* (STEC), enterpathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffuse adherent *E. coli* (DAEC) (Clements *et al.*, 2012; Nataro and Kaper 1998). Furthermore, the most important EHEC, which includes the –O111 and –O157 serogroups, results in majority of the reported severe cases of diarrhoea and haemorrhagic colitis (Okeke, 2009).
Faecal streptococci and enterococci	Faecal streptococci (FS) are defined as bacteria that is Gram-positive, catalase-negative, non-spore forming cocci that grow in medium containing bile salts and sodium azide agar at 35 °C, belonging to the genera *Enterococcus* and *Streptococcus* possessing group D Lancefield antigen (WHO, 2001). Enterococci are defined as all aerobic faecal streptococci that is able to grow at pH 9.6, 10 °C and 44 ± 1 °C in the presence of 6.5% sodium chloride. Nearly all FE are members belonging to the *Enterococcus* genus, and also fulfil the following criteria: resistance to 60°C for 30 min and ability to reduce 0.1% methylene blue (WHO, 2001). The enterococci are a subset of faecal streptococci that grow under the conditions outlined above, composing of the species *E. faecalis*, *E. faecium*, *E. durans* and *E. hirae*, which have been known for their association with faecal pollution (WHO, 2011). Members of this group of indicators are known to be present in large numbers in water contaminated by sewage, human and animal wastes.

1.3 Next-generation sequencing and water quality assessments

Genomics is a relatively new scientific discipline having nucleic acid sequencing as its core. Recent advances in the sequencing technologies has decreased in cost and scale over sequencing's 40-year history. This has allowed for massively parallel analysis of DNA sequence information generated from PCR amplicons or directly from nucleic acids. Nucleic acids could be isolated from simple single cells or complex communities. Hence, sequencing is said to have revolutionized all biological-related fields and ushered a new era of genomics, shifting the paradigm of genomics to address biological questions across the nucleotide to ecological scale (Koboldt *et al.*, 2013). In clinical research, next-generation sequencing (NGS) has been applied as a screening tool which could complement or even bypass conventional diagnostic methods of microscopy, Gram-staining and culturing for the detection and identification of etiological agents in disease, respectively (Kumar *et al.*, 2013). It has also been applied in the human microbiome profiling project (Qin *et al.*, 2010), sputum analysis of patients with cystic fibrosis (Lim *et al.*, 2014), viral pathogens detection (Yang, 2014) as well as many others. Similarly, NGS has also been applied for the assessment of various environmental scenarios, such as the determination of soil (Howe *et al.*, 2014) and river (Amos *et al.*, 2014) microbiota, communities found in activated sludge from WWTPs (Ju *et al.*, 2014), communities in sediments from an oil spill (Mason *et al.*, 2014), faecal microbiota (Ilmberger *et al.*, 2014) and diversity in aquatic settings (Doxey *et al.*, 2014). Overall, these studies have demonstrated the dynamic power of NGS and allowed for new opportunities to arise in water quality assessment through analysis of waterborne microbial communities. Analysis of waterborne microbial communities could advance development of new indicators or biomarkers required for microbial source tracking as well as novel understandings of important microbial mediated biological processes (Tan *et al.*, 2015a). As these new tools are being

developed, biases and uncertainties associated with nucleic acid-based methods is a topic which still needs to be critically considered and is therefore still at "work-in-progress" stage.

1.3.1 Technologies, experimental design and analysis methods for metagenomic studies

1.3.1.1 Next-generation sequencing technologies

Early studies investigating the relationship between water quality and waterborne microbial communities utilised the pyrosequencing platform from Roche 454 Life Sciences (McLellan *et al.*, 2010; Vandewalle *et al.*, 2012). The rationale behind the use was due to the relatively long sequence read lengths, initially at 110 bp but now up to 1 000 bp (van Dijk *et al.*, 2014a), as well as the overall optimized sequencing conditions and bioinformatic analysis workflows developed during that period (Sergeant *et al.*, 2012). However, as the sequencing by synthesis platform by Illumina was introduced into the market later on, with initial read lengths of 35 bp but increased later on from the merging of paired-end reads allowing up to 100 bp, the platform was able to rival pyrosequencing and saw increased standing in environmental sample analysis (van Dijk *et al.*, 2014a). Despite availability of other NGS platforms in the market, including sequencing by ligation (SOLiD) by Applied Biosystems, single molecule real-time sequencing (SMRT) by Pacific Biosciences and ion semiconductor sequencing by Ion Torrent. These platforms has been reported in studies which sequenced both PCR amplicons or direct nucleic acid (Marine *et al.*, 2014; Mitra *et al.*, 2013; Marshall *et al.*, 2012; Yergeau *et al.*, 2012), as well as their capability of generating 0.7 to 600 gigabases (Gb) of data efficiently and in a short amount of time. However, these platforms are not widely adopted for water analysis due to technical limitations in the sequence generated which still require further advancements. A comparison between the specifications of the first-generation and selected next-generation

sequencing methods and platforms are described in **Table 1.4**. Consequently, in the last few years, the Illumina MiSeq technology has become one of the most widely used sequencing platform in majority of published literature, and with the announcement by Roche to withdraw the GS FLX 454 pyrosequencing platform, current studies reported have embarked on further refining the data analyses stage tailored specifically for Illumina platforms by improving on methods in library preparation (Esling *et al*., 2015; Shishkin *et al*., 2015; Kozich *et al*., 2013;) and quality control of sequence reads (Schirmer *et al*., 2015; Nelson *et al*., 2014; Kozich *et al*., 2013).

1.3.1.2 Experimental design

Any experimental design should ultimately answer the scientific question of interest, though a reasonable and rigorous experimental design is important for acquiring high-quality and meaningful sequencing data. A typical design of metagenomic studies is summarized in **Figure 1.1**. Several reviews have identified key aspects that should be considered or at least recognized in terms of the strengths and pitfalls of current metagenomics-driven studies utilizing NGS (Ma *et al.*, 2014; Zhou *et al.*, 2014; Valverde and Mellado, 2013; Mande *et al.*, 2012; Prakash and Taylor, 2012; Gilbert and Dupont, 2011; Kunin *et al.*, 2008; Wooley, 2007). Firstly, the adequate number of biological and technical replicates of samples investigated is necessary for thorough and accurate statistical analyses. Many initial exploratory investigations reported were not able to appropriately address this issue due to the high cost of sequencing at that time as well as the recent realisation in the implications of biological and technical replicates in downstream analyses (Knight *et al.*, 2012). The importance of biological replicates from within an experimental group allows for a measure of within-group variation leading to a more accurate evaluation of the degrees and/or effectiveness of observed biological differences.

Technical replicates from a given sample provides a measure of the reproducibility of the sequencing data generated using identical molecular methods, such as DNA extraction and PCR amplification, and subsequent NGS protocols, such as library construction and sample multiplexing. Therefore, future studies should not be compromised in terms of appropriate replicates due to economic constraints as the cost of NGS is exponentially decreasing per year (Koboldt *et al.*, 2013; Muir *et al.*, 2016). Secondly, addition of positive and negative controls should be considered during the amplification steps, NGS as well as data analysis stages. This is to ensure that there is no contamination across samples and reagents, genomes are properly assembled from the analysis of the positive control and differential binning algorithms are implemented as intended. Thirdly, several critical decisions must be made at the sampling stages and include (i) the selection and monitoring of important influential parameters and their range of scales, such as physicochemical, biological, temporal and spatial variation; (ii) the methods for obtaining the above parameters, such as on-site or laboratory; (iii) the sample preservation and storage methods, such as on-site fixing or addition of preservatives; (iv) the methods of improving DNA or RNA quantity and purity; and (v) the sequencing platform selected, the depth and length of sequencing (Ju and Zhang, 2015).

Table 1.4: Comparisons between the specifications of first- and selected next-generation sequencing method and platforms† (adapted from Cheng *et al.*, 2013).

Sequencing platform	Read length (bp)#	Reads per run	Throughput (Gb)	Time per run	Accuracy (%)	Cost per 1 million bp	Advantages	Disadvantages
First-generation								
Sanger	400 – 900	n/a*	n/a*	20 min – 3 h	99.9	$$$$$	Long individual reads. Useful for many applications.	More expensive and impractical for larger sequencing projects. This method also requires the time consuming step of plasmid cloning or PCR.
Next-generation								
Illumina HiSeq™ 2000	35 – 100	300 Million – 3 Billion	100 – 600	2 – 11 days	98.0	$$	Ultra-high throughput. High capacity of multi-plexing.	Short read assembly may miss large structural variations. Signal interference among neighbouring clusters. Homopolymer errors.
Illumina Miseq™	35 – 150	1 – 25 Million	1.5	27 h	99.2	$	Well-proven sequencing technology. Fully automated workflows developed. Low cost. Fast runtime.	Low abundance of amplified template.
Applied Biosystems 5500 SOLiD™	35 – 75	1.2 – 1.4 Billion	120	6 – 14 days	99.9	$$	Ultra-high throughput. Two-base coding for higher accuracy. High capacity of multi-plexing	Short read assembly may miss large structural variations. Long run time. Signal interference among neighbouring clusters. Signal degradation.
Roche 454 Life Sciences GS FLX Titanium™	700 – 1000	1 Million	0.7	0.35 – 0.42 days	99.9	$$$$	Long read assembly allows detection of large structural variations. Short runtime.	Lower throughput. Homopolymer errors. Signal interference among neighbouring clusters.
Iron Torrent Ion Proton™ PI & PII	100 – 400	40 – 80 Million	10 – 30	4 h	98.5	$$$	Fast runtime. Highly scalable due to different chips available. Low cost.	Newest to the market
Iron Torrent Ion PGM™ 314, 316 & 318 chips	100 – 400		0.01 – 1	1 – 3 h	98.5	$$$	Highly scalable due to different chips available. Low cost. Fast run time.	Homopolymer errors.

*n/a, not applicable
#average read length depends on specific sample and genomic characteristics.
†specifications for all platforms are derived from company websites.

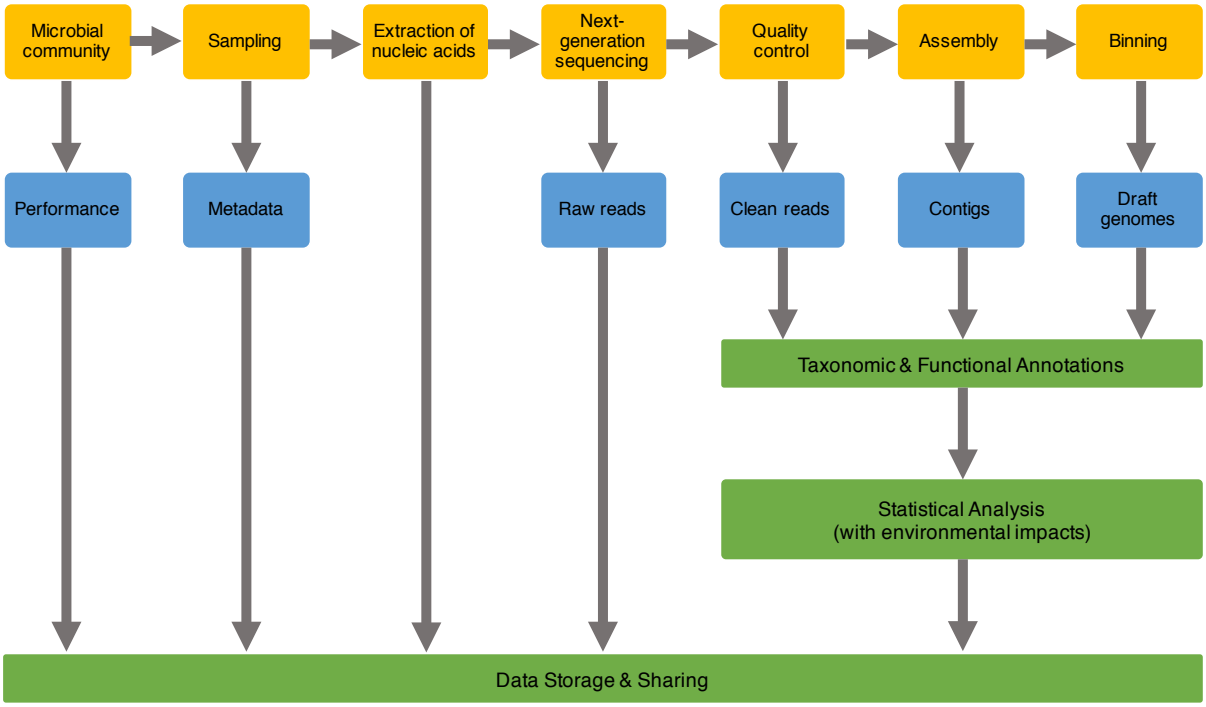


Figure 1.1: Flowchart of experimental preparation and data analysis in a typical metagenomic study; some typical bioinformatics tools or platforms are listed above the arrow lines (adapted from Ju and Zhang, 2015).

1.3.1.3 Methods for analysis

Analysis of metagenomic data is performed by the relatively new scientific field of bioinformatics, where various routes are available depending on the scientific question as summarized in **Table 1.5**. For example, short reads could be assembled into contigs and subsequently reconstructed into draft genomes through what is known as a self-accelerating data mining circle as depicted in **Figure 1.2** (Ju and Zhang, 2015). The operation of this circle on NGS data annotation allows for novel training sets or generation of reference sequencing for the next rounds of reads, contigs, and genome annotations. This ultimately enables the discovery of novel conditions, such as biochemical attributes or metal resistance, for the

selective enrichment and isolation of novel uncultured microbes. Successful isolation of these novel microbes will validate the predicted annotations, increase biochemical properties and function as well as retrieval of their complete genomes, thereby increasing knowledge on the novel organism. Hence, this approach provides a powerful way for identifying novel microbial resources which may be important in biotechnological fields (Ju and Zhang, 2015).

Figure 1.2: Schematic circle of self-accelerating data mining from raw reads, contigs, and genomes in a metagenomics-based study. Reads provide an overview of microbial community structure and functions; Contigs provide an overview of gene catalogue for the discovery of novel resources and patterns; Genomes provide a window to the function and conditions for isolation of uncultured microorganisms; Isolates provide function verification of genomes, physio-biochemical attributes, and materials for future application in genetic engineering (adapted from Ju and Zhang, 2015).

Table 1.5: Platforms and software tools available for the bioinformatics analysis of metagenomes (adapted from Ju and Zhang, 2015).

Analysis Steps	Bioinformatic Tool	Tool Application	Reference
Quality Control	MG-RAST	de-multiplexing of barcoded samples; quality control; dereplication	(Meyer *et al.*, 2008)
	CLC bio	adapters/linkers removal; de-multiplexing of barcoded samples; quality control; dereplication; overlapping of paired-end reads	http://www.clcbio.com/
	IMG/M system	quality control; dereplication	(Markowitz *et al.*, 2012)
	PRINSEQ	quality control; dereplication	(Schmieder and Edwards, 2011a)
	NGS QC Toolkit	adapters/linkers removal; quality control;	(Patel and Jain ,2012)
	DeconSeq	DNA contamination removal	(Schmieder and Edwards, 2011b)
	FASTX-Toolkit	adapters/linkers removal; de-multiplexing of barcoded samples; quality control	http://cancan.cshl.edu/tools.php
Assembly	Velvet	genome assembler	(Zerbino and Birney, 2008)
	ABySS	contig/genome assembler	(Simpson *et al.*, 2009)
	SOAPdenovo2	genome assembler	(Xie *et al.*, 2014)
	CLC bio	genome/metagenome assembler	http://www.clcbio.com/
	IDBA-UD	metagenome assembler	(Peng *et al.*, 2012)
	MetaVelvet	metagenome assembler	(Namiki *et al.*, 2012)
	Ray Meta	metagenome assembler	(Boisvert *et al.*, 2012)
	Omega	metagenome assembler	(Haider *et al.*, 2014)
	MEGAHIT	metagenome assembler	(Li *et al.*, 2015d)
Binning	GroopM	genome reconstruction	(Imelfort *et al.*, 2014)
	CONCOCT	composition-based taxonomic binning/assignment	(Alneberg *et al.*, 2014)
	MaxBin	composition-based taxonomic binning/assignment	(Wu *et al.*, 2014)
	METABAT	composition-based taxonomic binning/assignment	(Kang *et al.*, 2015)
	PhyloPythiaS	composition-based taxonomic binning/assignment	(McHardy *et al.*, 2007)
	TETRA	composition-based taxonomic binning/assignment	(Teeling *et al.*, 2004)
	CompostBin	composition-based taxonomic binning/assignment	(Chatterji *et al.*, 2008)
	TACAO	composition-based taxonomic binning/assignment	(Diaz *et al.*, 2009)
	MetaPhlAn2	homology-based taxonomic binning/assignment	(Segata *et al.*, 2012)
	MetaPhyler	composition & homology-based taxonomic binning/assignment	(Liu *et al.*, 2010)
	PhymmBL		(Wang *et al.*, 2012b)
	MetaCluster		(Wang *et al.*, 2012c)
Annotation	MG-RAST	reads annotation	(Meyer *et al.*, 2008)
	IMG/ER	contigs and genomes annotation	(Markowitz *et al.*, 2009)
	RAST	contigs and genomes annotation	(Aziz *et al.*, 2008; Overbeek *et al.*, 2014)
	WebMGA	reads and contigs annotation	(Wu *et al.*, 2011)
	ggKbase	contigs and genomes annotation	http://ggkbase.berkeley.edu/

1.3.1.3.1 Quality control

Regardless of the route taken for the analysis of NGS data, quality control of raw reads obtained from NGS platforms is an essential step to guarantee the use of high quality sequencing (also known as clean reads) in downstream analysis. Typically, quality control includes (i) quick summary statistics of raw data quality of a NGS library. This is achieved by the use of base quality, ambiguous bases, sequence duplication levels, length distribution, GC content and adaptor content which could be obtained from open-source tools such as FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, (ii) nucleic acid adaptors or linkers used in NGS needs to be removed, followed by (iii) de-multiplexing, (iv) overall quality check of the generated reads, and then (v) dereplication, which is the removal of unwanted duplicate or near-duplicate artefacts produced during NGS library preparations. This is commonly seen with large insert size libraries. Failure in the dereplication step of quality control has been reported to lead to biased measures of taxon abundance or gene expressions (Schmieder and Edwards, 2011a). To date, there is no uniform standard for the dereplication of metagenomes. For example, the IMG/M annotation pipeline only retained one copy of those DNA sequences with greater than 95% similarity (Markowitz *et al.*, 2014). Whilst the MG-RAST pipeline removes all but a single representative of the clusters of DNA sequences whose first 50 bp are identical (Meyer *et al.*, 2008). Conversely, the PRINSEQ offers the complete list of options for users to remove 100% similarities and reverse complement duplicates (Schmieder and Edwards, 2011a). Finally, (vi) reads overlapping can be constructed for paired-end (PE) library NGS. Paired-end sequencing allows for overlapping and merging of the PE sequences obtained. Advances of using overlapped sequences with extended length means that higher resolution and accuracy of taxonomic and functional annotation would be possible due to the higher quality and smaller error of obtained sequences (Masella *et al.*, 2012; Schloss *et al.*, 2009). Popular bioinformatic tools for overlapping PE sequences include QIIME

(Caporaso *et al.*, 2010), Mothur (Schloss *et al.*, 2009), FLASH (Magoč and Salzberg, 2011) or PANDAseq (Masella *et al.*, 2012) to name a few. Generally, popular packages or automated workflows have been designed for processing the raw NGS data, such as MEGAN, IMG/M and MG-RAST for DNA analysis or PRADA for RNA analysis (Torres-García *et al.*, 2014). Among those listed in **Table 1.5**, PRINSEQ and NGS QC Toolkit have been the most popular tools used for DNA sequences, whilst RSeQC (Wang *et al.*, 2012) is the most popular for RNA sequences followed by RNA-SeQC (DeLuca *et al.*, 2012).

1.3.1.3.2 Assembly

Assembly is the computational process of connecting reads which passed quality control to yield long contigs. This will enable for the prediction of open reading frames, protein coding genes transcripts, strain specific genomic islands or allow for recovering of genomic sequences (discussed in the following section). This in turn, will allow for a qualitative analysis of the genetic content of the generated sequences at various taxonomic levels, and is particularly important for unculturable microorganisms. Although requiring substantial computational resources, metagenomic assembly effectively reduces data sizes (Howe *et al.*, 2014). However, it has been suggested that this step is avoided in studies investigating biospheres with low sequencing depth or coverage due to the introduction of potential bias from the suppression of low abundance species which is omitted during assembly of reads (Rosseel *et al.*, 2013; Sims *et al.*, 2014). Typically, two strategies of assembly are employed depending on the scientific question and availability of other sequences, either *de novo* or reference based. *De novo* based assembly is usually recommended and widely adopted to avoid discarding the novel sequences described above regardless of the computationally expensive resources required, which previously reached up to hundreds of gigabytes of RAM alone with weeks of computational

runs. However, recent advances in the computational fields have greatly reduced the resource required (Howe *et al*., 2014). Most *de novo* assembly makes use of a mathematical concept known as the de Bruijn graph, and is implemented in tools such as MetaVelvet (Namiki *et al.*, 2012), SOAPdenovo2 (Xie *et al.*, 2014), IDBA-UD (Peng *et al.*, 2012). These three tools have been widely adopted in published literature utilising the *de novo* assembly approach. IDBA-UD and Meta-IDBA both adopt a multiple k-mer strategy and have been demonstrated to be far more memory-efficient, only requiring nearly up to 400% less memory than Velvet, MetaVelvet, and SOAPdenovo (Namiki *et al.*, 2012; Peng *et al*., 2012). MetaVelvet has been reported to provide significantly longer scaffold than Velvet, SOAPdenovo, and Meta-IDBA and increases the number of predicted genes from microbial metagenomic data sets of the human gut (Namiki *et al.*, 2012). Additionally, comparison between MetaVelvet, SOAPdenovo2 and CLC bio's *de novo* tools in a study showed that CLC bio's *de novo* assembler produced longer scaffolds, based on the N_{50} size, mean length and maximum length of phenol-degrading methanogenic metagenome, in addition to the faster time and smaller memory required (Ju and Zhang, 2014). Reference based assembly utilises reference sequences and aligns reads with similar homology accordingly. Bioinformatic tools which utilise this strategy, such as Roche's Newbler, are fast and memory efficient (Kunin *et al.*, 2008). Though, it should be noted that reference based strategy is limited in its ability to truly capture the differences between genomes of novel species due to high genetic novelty and diversity and thus greatly underestimates the microbial diversity in complex microbial systems such as soil or water (Howe *et al.*, 2014).

1.3.1.3.3 Binning

Shotgun sequencing allows for bypassing of the cultivation limitation due to direct sequencing of isolated DNA. Furthermore, it allows for discovery of novel organisms and genes with recovery of near-complete genome sequences of species which could allow for further metabolic, evolutionary and ecological understandings (Albertsen *et al.*, 2013; Sharon *et al.*, 2012; Wrighton *et al.*, 2012). Binning is an important step and refers to the computational process of clustering contigs or reads into separate groups that might represent an individual level of taxonomic resolution originating from the same population source, such as single genome, similar genomes from the same genus or genomes from the same taxa (Strous *et al.*, 2012). Metagenomic binning methods can be based on various factors, *viz.* homology, abundance or coverage, sequence composition, or a combination of these factors. Homology-based tools are typically used for taxonomic classification of shotgun metagenomic reads based on similarity comparisons with a database of known reference marker genes, such as the 16S rRNA genes, *rpoB* gene, or specific markers unique to a species. As homology binning strategy utilises similarity searchers and taxonomic assignments, they are not reliable for assigning short reads and often require longer assembled contigs and manual efforts to ensure high assignment accuracies, with each tool having their own minimum contig length requirements. Though this approach may provide useful grouping information on the short reads generated for improving downstream genome binning of specific organisms, the approach is limited by the heavy reliance on the quality and representativeness of the reference databases used, poor taxonomic resolution, accuracy as well as sensitivity of the alignment tool (Alneberg *et al.*, 2014; Sharpton, 2014). Tools such as MetaPhlAn2 (Segata *et al.*, 2012), MetaPhyler (Liu *et al.*, 2010) and CARMA3 (Gerlach and Stoye, 2011) fall into this category within the toolbox. Abundance or coverage-based tools classifies contigs by their differential coverage profiles among multiple related metagenomes. This is done with the assumption that contigs from the same

microorganisms should have similar abundance profiles in any given metagenome. The draft genomes obtained from this method can then be further purified and refined using composition-based methods and other strategies, including reassembly, manual curation and paired-end tracking (Albertsen *et al.*, 2013; Karlsson *et al.*, 2013; Sharon *et al.*, 2012). Composition-based algorithms generally use conserved nucleotide compositions derived from NGS data, such as GC content of DNA sequences within genomes or tetranucleotide frequencies (TNF) to group different contigs or reads. Typically, composition based tools are improved by providing sample-specific training datasets. To date, two unsupervised approaches have been widely applied to reconstruct high-quality genomes of uncultured organisms directly from metagenomes. The first approach from the Tetra-ESOM tool (Dick *et al.*, 2009) explores genome signatures by clustering tetranucleotide frequencies using emergent self-organizing maps (ESOM). This new composition-based method has been successfully utilized to retrieve numerous genomes from the metagenomes of the Guaymas Basin hydrothermal plume and surrounding waters (Sheik *et al.*, 2014), acidophilic biofilm communities (Dick *et al.*, 2009) and acetate-amended aquifer sediment communities (Wrighton *et al.*, 2012) to name a few. Tools such as TETRA (Teeling *et al.*, 2004), PhyloPythia (McHardy *et al.*, 2007), CompostBin (Chatterji *et al.*, 2008) and TACAO (Diaz *et al.*, 2009) fall within this category within the toolbox. On the other hand, bioinformatic tools such as PhymmBL (Brady and Salzberg, 2009) and MetaCluster (Wang *et al.*, 2012b) both utilise the composition and homology approach in taxonomic classification or clustering of reads from the metagenomic sequences into same/or similar genomes. Recently, tools including GroopM (Imelfort *et al.*, 2014), CONCOCT (Alneberg *et al.*, 2014), MaxBin (Wu *et al.*, 2014) and METABAT (Kang *et al.*, 2015) were developed to integrate coverage/abundance profiles and TNF patterns of contigs across multiple temporal or spatial related metagenomes into efficient automated pipelines for genome reconstruction from metagenomes. The completeness and potential contamination in

reconstructed genomes have been estimated by the presence/ absence of essential marker genes, such as essential single copy marker genes conserved in 95% of bacteria (Dupont *et al.*, 2012), conserved phylogenetic marker genes (Wrighton *et al.*, 2012), or clusters of orthologous groups (COGs) (Raes *et al.*, 2007). Currently, the recently developed CheckM is the only automated tool that can assess the quality of a genome recovered from isolates, single cells, and metagenomes based on these conserved marker genes (Parks *et al.*, 2015).

1.3.1.3.4 Annotation

Both assembled contigs and unassembled reads from NGS data could be used for annotation, thereby identifying the sequence and inferring its functions. Generally, assembled contigs with more compact size and much longer lengths allow for rapid analysis of both specific species and their functional genes compared with unassembled short reads. However, an assembly based annotation approach has the potential to introduce biases for quantitative analysis due to the difficulty associated with low abundance species and closely related strain as well as the exclusion of unassembled data in downstream analysis. In contrast, unassembled short reads retain all the original abundance information and enable quantitative comparisons within or between habitats or ecosystems. However, short reads are large in terms of data size and may lack resolution for taxonomic and functional annotations (Li *et al.*, 2015a; Wang *et al.*, 2012). For assembled contigs, gene calling tools commonly used to determine protein encoding genes from predicted open reading frames include FragGeneScan (Rho *et al.*, 2010), Orphelia (Hoff *et al.*, 2009), MetaGeneMark (Zhu *et al.*, 2010), MetaGeneAnnotator (Noguchi *et al.*, 2008) and Prodigal (Hyatt *et al.*, 2010), with the latter being the popular tool due to lower false positive rates and improved gene structure predictions. Quantification of predicted protein encoding genes from assembled contigs typically is based on the mapping of the contigs to

known genes with Reads Per Kilobase of transcript per Million mapped reads (RPKM) (Mortazavi *et al.*, 2008), Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) (Trapnell *et al.*, 2010) and relative abundance (number of reads mapped divided by total number of reads) metrics widely used. RPKM has been proposed to better reflect the absolute transcriptional activity due to the potential of uneven richness of different protein encoding genes in a given sample (Xia *et al.*, 2014). Popular alignment tools for mapping of the contigs to known genes is Bowtie2 (Langmead and Salzberg, 2012; Langmead *et al.*, 2009), BWA/BWA-SW (Li and Durbin, 2010, 2009) and recently developed HISAT (Kim *et al.*, 2015). Alternatively, binning approaches described above could be used to recover draft genomes of uncultured organisms and submitted to web-based servers with established pipelines such as IMG/ER (Markowitz *et al.*, 2009), RAST (Aziz *et al.*, 2008), WebMGA (Wu *et al.*, 2011) and ggKbase. For unassembled reads, the the direct use of clean unassembled reads for quantitative analysis is typically first approach. Additionally, genomic and transcriptomic data which are available from the same samples allow for understandings of the active microbes as well as enable comparisons in their expression patterns (Frias-Lopez *et al.*, 2008; Yu and Zhang, 2012). For example, unassembled reads have been used to characterize or predict environmental biohazards such as bacterial pathogens, viruses and antibiotic resistance genes (ARGs) in drinking water disinfection systems (Gomez-Alvarez *et al.*, 2012; Li *et al.*, 2015c) and waste and wastewater treatment systems (Li *et al.*, 2015b; Bibby and Peccia, 2013; Yang *et al.*, 2013; Tamaki *et al.*, 2012). Taxonomic and functional annotation with unassembled reads is based on homology searches against a reference database using various alignment tools either *via* standardised platforms or established pipelines such as MG-RAST for DNA fragment analysis and PRADA for RNA analysis. Taxonomy profiling tools, such as Pathoscope (Francis *et al.*, 2013), Sigma (Ahn *et al.*, 2015), and MetaPhlAn2 (Segata *et al.*, 2012), provide species level resolution for performing bio-surveillance and detecting

biohazards in healthy human microbiome with the assistance of reference genomes (Consortium, 2012). Local similarity searches with BLASTX or PSI-BLAST is time consuming and computationally expensive given the large sizes of reference databases and query data. To cope with this limitation, several tools have been developed and shown to be much faster in speed but at the cost of reduced levels of sensitivity and specificity. Tools such as BLAT (Kent, 2002), RapSearch2 (Zhao *et al.*, 2012), USEARCH (Edgar, 2010) and DIAMOND (Buchfink *et al.*, 2014) fall within this category in the toolbox. To maximize functional annotation, recent versions of popular automated pipelines such as MG-RAST, IMG/M and MEGAN have merged majority of the publically available databases (KEGG, UniRef, Swiss-Prot, PFAM, TIGRFAM, eggNOG, COG, CDD and SMART) into a single framework to allow for for user-friendly and streamlined analyses (Huson *et al.*, 2016; Thomas *et al.*, 2012).

1.3.1.3.5 Data storage and sharing

Sharing of NGS data, analysis results and associated metadata is a traditional and efficient method of knowledge dissemination. Its significance lies in the beneficial outcomes of comparative studies and complete elimination of unnecessary repeated sequencing or analysis of similar microbial systems. Several publically available databases have been maintained to promote the storage and sharing of NGS data, such as NCBI-SRA, MG-RAST and GOLD. Submitted data typically are required to be coupled with completed metadata file that is prepared according to the recently established minimum information about any sequence checklist (MIxS) from the Genomic Standards Consortium (http://gensc.org/projects/mixs-gsc-project/) and include information such as sampling habitat type, location, time, organisms, sequencing method, analysis method (Ju and Zhang, 2015).

1.3.2 Applications of next-generation sequencing in water quality analysis

1.3.2.1 Microbial quantification

The application of NGS in water quality analysis is still in its early stages and has not been integrated into established epidemiological framework, whereby observed trends with NGS studies is correlated to public health risks. This is due to the fact that a fundamental requirement of microbial risk assessment within the quantitative microbial risk assessment (QMRA) framework is to be able to accurately quantify biological agents which leads to health risks (Haas *et al.*, 2014). Several molecular based methods are known to be highly quantitative and is able to provide results in the form of absolute gene copy number *viz.* quantitative-PCR (qPCR) and digital-PCR (dPCR). On the other hand, the profiles of species generated from NGS amplicon and shotgun sequencing is regarded more as being qualitative result than absolute quantification. Furthermore, as amplicon sequencing is largely based on PCR amplification, the method is subjected to limitations associated with PCR such as generation of false positives, GC content of the amplicons, secondary structure formation as well as the impact of primers targeting the different SSU rRNA regions on resultant amplicons. These factors could lead to under-estimations or over-estimations of relative abundance and diversity in downstream analysis from the subsequent sequencing error or chimera formations. Chimera formations refers to incorrectly assembled amplicons and can produce artificial sequences that are often difficult to identify (Schirmer *et al.*, 2015; Nelson *et al.*, 2014; van Dijk *et al.*, 2014b; Kozich *et al.*, 2013; Quail *et al.*, 2012; Wylie *et al.*, 2012 Kozarewa *et al.*, 2009). This is further complicated by recent studies which reported pronounced biases associated with increasing PCR cycle numbers (Murray *et al.*, 2015; Schirmer *et al.*, 2015; Sinclair *et al.*, 2015) well as the fact that some bacterial species house multiple copies of the SSU rRNA gene (Angly *et al.*, 2014). Additionally, a study reported sequencing errors associated with reads generated from

the Illumina MiSeq platform may also result in differences due to differences in the library preparation protocols used (Schirmer *et al.*, 2015). Studies which utilised mock NGS data of bacterial communities suggest that these variations can sometimes result in overestimation of the number of OTU observed and therefore skew estimates of species richness and evenness (Kennedy *et al.*, 2014; Nelson *et al.*, 2014). Hence, all of the factors described above influence the relative abundance of the OTU observed in downstream bioinformatic analysis and could lead to misinterpretations of the results observed. Despite the concerns discussed above, several studies have demonstrated usefulness of NGS targeting the SSU rRNA gene in determining the microbial species composition and quantification of their relative abundance in water samples (Ong *et al.*, 2013; Wang *et al.*, 2014). Although, the accuracy of these studies should be interpreted with caution as it is likely sample dependent and may be influenced by the factors described above. Therefore, as there are inherent differences between any sample types and its associated microbiome, it is likely that there could be no universal approach for quantitative measurement of all sample types and is a major limitation of this approach (Murray *et al.*, 2015; Schmidt *et al.*, 2015).

In its applications, amplicon sequencing typically only provides insight into the taxonomic composition of the sample. It is impossible to directly resolve the biological functions associated with these taxa using this approach. In some cases, phylogenetic reconstruction can be used to infer biological functions that are encoded in a well annotated genome containing the particular SSU rRNA sequence observed. But, the accuracy with which these methods estimate the true functional diversity of a community is limited due to the fact that only few genomes are well understood, annotated and publically available in sequence databases (Langille *et al.*, 2013). Due to these limitations as well as the concerns regarding introduction

of biases in PCR-based NGS studies, several studies have investigated microbial communities with an amplification-free approach targeting the SSU rRNA genes from shotgun NGS data. The shotgun sequencing approach utilises extracted nucleic acids from all cells in a sample. Rather than specific targeting of a set of sequences or gene for amplification, all nucleic acids are fragmented/sheared into tiny segments that is subsequently sequenced on a NGS platform (Sharpton, 2014). For example, Logares *et al.* (2013) recently assessed the diversity of marine plankton using three different NGS platforms with this approach as well as the amplicon approach, *viz.* shotgun sequencing by Illumina HiSeq and Roche 454 as well as amplicon sequencing of targeted SSU rRNA gene by Roche 454. The composition and diversity of the SSU rRNA genes observed from shotgun sequencing by Illumina was higher and more even than shotgun and amplicon sequencing by the Roche 454 platform. Furthermore, the relative abundance of the microbial taxa observed from shotgun Illumina sequencing was comparable to relative abundances observed by catalysed reporter deposition fluorescence *in situ* hybridization as well as flow cytometry assays. Hence, the choice of NGS platform may yield different results due to varying resolution and quantitative power from sequence throughput, coverage and depth. By spiking samples with known standards prior to DNA or RNA isolation, several protocols have been developed and reported in a few studies with the intention of quantifying gene copy numbers or transcripts from shotgun NGS data (Gifford *et al.*, 2011; Satinsky *et al.*, 2014; Rivers et al., 2013). For example, use of these controls for benchmarking was applied in a studies investigating microbial communities in the Amazon river by (Satinsky *et al.*, 2015, 2014) where DNA and RNA standards of known concentrations which were prepared by using commercially acquired plasmid DNA. The plasmid was linearized and subsequently subjected to restriction enzyme digestion. For RNA quantification, the linear plasmid DNA fragments were further transcribed *in vitro* and the resultant product's absolute copy number quantified *via* spectrophotometry. Thereafter, both DNA and RNA standards was

added to the samples of the study prior to cell lysis, total metagenome and metatranscriptome isolation and NGS. The copy number of the sequenced internal standard was then used as a benchmark to estimate the sequencing depth and allow for relative quantification of the differential expressed transcripts of the studied samples.

Though potential biases and uncertainties introduced by nucleic acid-based methods and NGS still limit these two approaches, the studies discussed above demonstrate the potential use of NGS for quantification of either the relative abundance of targeted genes or taxonomic groups or determining the absolute quantification through the addition of control standards to environmental samples. However, as these methodologies is still relatively new with unforeseen pitfalls, developments and results obtained from further investigations should always be carefully interpreted.

1.3.2.2 Faecal and pathogen detections

Contamination of freshwater originally intended to serve as sources for drinking water or domestic, agricultural and industrial purposes continues to be a cause for concern. To determine the source of contamination of these resources, molecular based methods have been a favourable approach. NGS of the SSU rRNA gene has been used to characterize the microbial population in raw sewage/wastewaters entering wastewater treatment plants in several countries across the globe. Overall, bacterial communities associated with wastewater infrastructures has been observed to be unique and different from those present in human faecal materials as well as other environments such as soil and the ocean (Cai *et al.*, 2014; Shanks *et al.*, 2013; McLellan *et al.*, 2010; Unno *et al.*, 2010). Specifically, comparisons between the

wastewater microbial communities and human faecal communities from the Human Microbiome Project suggested that there is a difference of approximately 90 to 95% between the two sources, suggesting that only approximately 10% to 15% of the microbial populations in wastewaters were of human origin. Additionally, the studies reported the Firmicutes and Bacteroidetes genera to dominate the bacterial population from the percentage of detected faecal microbes, followed by other anaerobic families such as Bifidobacteriaceae, Coriobacteriaceae. Another interesting study reported the statistical correlation of the dominant families of Bacteroidaceae, Prevotellaceae, Lachnospiraceae and Ruminococcaceae to the obesity rate of the cities investigated, where higher obesity rates corresponded to increased relative abundance of Bacteroidaceae in domestic wastewaters. This finding is in line with a previous finding suggesting increased representation of Bacteroidaceae in human gut microbiomes of high fat consumption diets (David *et al.*, 2014). As the major proportion of the microbial communities in wastewaters were reported to not be of faecal origins, this population was observed to vary in diversity and abundance between different geographical locations and seemed to be unique in the taxa determined compared to animals, freshwaters and other environmental sources, on top of having an overall higher diversity and species richness compared to stool samples (Newton *et al.*, 2015; Shanks *et al.*, 2013; Vandewalle *et al.*, 2012; McLellan *et al.*, 2010). Authors of these studies have proposed the use of species which have been found in human faeces and in wastewaters, but absent or in low abundance in freshwaters to be new indicators for sewage contamination in water contamination studies. Although, several studies have emphasised the use of species or bacterial taxonomic groups identified, NGS approaches are site specific as different wastewater infrastructures, urban population, climate variations, rain and storm water infiltrations contribute to different microbial compositions (McLellan *et al.*, 2010; Shanks *et al.*, 2013; Ye and Zhang, 2013)

While these studies have shown the applicability of NGS in determining faecal contamination and revealing the microbiome present in contaminated waters, a direct approach to determine pathogen diversity and their abundance is also possible with NGS. Though traditional determinations of pathogens are costly, time consuming, technically challenging and requires prior knowledge on the targeted pathogens (Varela and Manaia, 2013), use of NGS overcomes these limitations and allows for screening of sequences with high similarity to waterborne pathogens, on top of potential quantification inferences in their relative abundance (Lu *et al.*, 2015; Cai *et al.*, 2014; Cai and Zhang, 2013; Ibekwe *et al.*, 2013; McLellan *et al.*, 2010). Identification of pathogenic sequences in early studies were not possible due to the short read lengths which could not provide the taxonomic resolution beyond genus or family levels. However, with increases in the NGS read lengths over the past few years, classification with higher confidence to the species level has now become possible. For example, a recent study investigating the bacterial pathogens in wastewater treatment plants utilised a combination of NGS and qPCR to determine the occurrence of bacterial pathogens at various treatment stages of the wastewater treatment process (Lu *et al.*, 2015). The study showed that the raw influent received by the plant was high in the levels of potential pathogens which were related to *Arcobacter butzleri*, *Klebsiella pneumonia*, and *Aeromonas hydrophila*, with the *Arcobacter* genus accounting for as high as 97.37% of the pathogenic populations across the various treatment stages. However, upon treatment, the wastewater effluent only contained *Arcobacter butzleri*, demonstrating the effectiveness of the biological process in removal of these pathogens. Furthermore, quantification by qPCR saw similar trends observed by NGS and confirmed these observations. Similarly, (Ye and Zhang, 2011) investigated wastewater treatment plants in four different countries (USA, China, Singapore and Canada) utilising the 454 pyrosequencing platform and reported the presence of pathogenic bacterial population in the raw wastewater influent received by the plants, with dominance corresponding to the

Aeromonas and *Clostridium* genera. The study saw dominance of *Aeromonas veronii*, *Aeromonas hydrophila*, *Clostridium perfringens* and *Corynebacterium diphtheria* species.

Recent studies have also utilised NGS in combination with first-generation sequencing and conventional approaches such as PCR and qPCR, in order to identify and quantify pathogens associated with outbreaks of disease. For example, this approach was used for the analysis of eight pathogenic viruses (norovirus, astrovirus, rotavirus, adenovirus, Aichi virus, parechovirus, hepatitis A virus, and hepatitis E virus) from wastewaters in Gothenburg, Sweden and prevented a potential outbreak of Hepatitis A and norovirus due to their early detections with two strains of hepatitis A virus involved in an ongoing outbreak in Scandinavia (Hellmér *et al.*, 2014). In another study, this approach was also used to understand the spread and distribution of the polio virus as well as differentiate wild-type and vaccine strains in Congo through the 2004 to 2011 years. The study was able to identify at least seven circulating lineages, determine their corresponding geographic regions in the country and show multiple independent emergence of the virus during the seven year period (Gumede *et al.*, 2013).

Together, the above studies have demonstrated the use of NGS in determining faecal, potential pathogen and viral contamination in water sources. However, the approach has not been integrated into the QMRA framework described above. On top of the issue with absolute quantifications described above, pathogen detection at a species taxonomic resolution has only been recently possible due to the increase in sequencing reads. Furthermore, though species resolution is now possible, the issue with strain specific virulence has not been overcome and is a topic for future research. Nonetheless, these studies have shown the potential impact of

NGS method providing insights into the distribution and relative representation of pathogens in different environmental scenarios. Such information with further developments and research should allow for better practices to safeguard public health. Therefore, one should see the integration of NGS into routine water quality analysis in the near future.

1.3.2.3 Chemical pollution

Chemical pollution is one of the major causes of diminished water quality. Several strategies for effective pollutant treatment have been through bioremediation by the stimulation of indigenous microbial communities for removal, immobilization or transformation of the compounds (Arjoon *et al.*, 2015; Gieg *et al.*, 2014; Koenig *et al.*, 2014; Arjoon *et al.*, 2013; Major *et al.*, 2002). As the presence of these chemical pollutants provides selective pressures which enrich for specific microbial populations, some of these microbes have been observed to be capable of coping with associated stresses in addition to utilising the contaminants as carbon or nitrogen sources, or even as electron acceptors for their respiration (Hemme *et al.*, 2010; Smith *et al.*, 2012). Analysis of enriched microbial communities as well as their genes in polluted waters through application of NGS has revealed key microbial processes involved in contaminant tolerance and transformation.

Several investigations (Tan *et al.*, 2015; Engel and Gupta, 2014; Rivers *et al.*, 2013; Mason *et al.*, 2012; Hemme *et al.*, 2010) of marine samples demonstrated the enrichment of microbial genes in unique environments and reported their involvement in these key biogeochemical processes. These studies examined the gene expression profiles of microbial communities from polluted waters and was able to determine the pathways taken by microbes during

transformation of the contaminants. The observed pathway profiles served as indicators of *in situ* biodegradation processes and allowed for exploitation to further accelerate the bioremediation process by shifting processes towards an intended pathway (Coleman and Chisholm, 2010; Kelly *et al.*, 2013; Ulloa *et al.*, 2012).

In a study by Hemme *et al.* (2010), microbial communities determined through NGS coupled with metagenomics was observed with limited diversity, consisting of beta- and gamma-proteobacteria, in a groundwater resource heavily contaminated with nitric acid, organic solvents and heavy metals for approximately 50 years. The study was able to confidently reconstruct putative community metabolisms of the beta- and gamma-proteobacterial species as well as propose possible lateral gene transfer within the community as a rapid key functional response and adaptation to the environmental contamination upon sequence evolutionary analysis. A recent study was also able to show the link between gene content and metabolic pathways taken by microbial communities contaminated with hydrocarbons (Tan *et al*., 2015a). Here, the authors were able to utilise SSU rRNA pyrosequencing and shotgun sequencing in order to determine whether differences in three hydrocarbon degrading methanogenic cultures established from two geographically distinct environments and incubated with different hydrocarbon substrates affected the genetic potential and composition of the microbial communities. Despite differing hydrocarbon substrates and inoculum sources, all three cultures showed functionally redundant genes with multiple features associated with syntrophic hydrocarbon conversion to methane. Other recent studies which surveyed the open oceans and coastal shorelines in Gulf of Mexico after the Deep Horizon Oil spill in 2010 utilising NGS reported microbial community consisting of mainly marine gamma-proteobacteria dominated the oil plumes within a water column. Furthermore, members of the class were determined to

be key players in the aerobic degradation of oil and gas contaminants. Analysis of the SSU rRNA genes from the microbial communities in beach areas before and after the spill was reported to shift from predominantly enteric species, typically associated with anthropogenic sources, to marine species associated with remediation efforts, such as Rhodobacterales, Oceanospirillales, and Rhodospirillales taxonomic orders. Furthermore, potential populations with capability to assist in remediation was also identified (Engel and Gupta, 2014; Rivers *et al.*, 2013; Mason *et al.*, 2012).

Full-scale wastewater treatment plant studies utilising NGS have reported the need for more reference genomes of species from activated sludge involved in key processes. Availability of these reference genomes will allow for interpretation of *in situ* studies and in this context, NGS data could then be fruitfully employed to infer media formulation and growth conditions for organisms of interest (Albertsen *et al.*, 2012; Kristiansen *et al.*, 2013; McIlroy *et al.*, 2013). While single omic studies have greatly enhanced our understanding and provided novel insights into the functional capabilities of the microbial communities of interest, integrated omics over space and time, which is any possible combination of meta- genomics, -transcriptomics and -proteomics have been demonstrated to further our understanding on the genes involved, actual expression and action of key processes in treatment of contaminated water sources under different environmental conditions such as alternating anaerobic/aerobic phases (Haroon *et al.*, 2013; Yu and Zhang, 2012). Furthermore, these in combination with physico-chemical parameters will allow for reconstruction of the ecological networks and detailed definition of organismal niches which may then be utilised to identify overall microbial community structure and function, which in turn could ultimately be harnessed for comprehensive reclamation of energy and biotechnologically relevant products from

wastewater, leading to a more sustainable treatment of wastewater resources (Muller *et al.*, 2013).

1.3.2.4 Detection of antibiotic resistance determinants

Antibiotics are at the centre of modern clinical medicine with their successes in reducing childhood mortality and increased overall life expectancy. They are crucial for invasive surgery and treatments such as chemotherapy but the number of infections caused by multidrug-resistant bacteria (MRB) is increasing across the world with the spectre of untreatable infections steadily becoming a reality. The most recent World Economic Forum Global Risks reports as well as the WHO have both indicated the impact of antibiotic resistance (AR) and stated AR as one of the greatest threats to human health in the 21st century (Blair *et al.*, 2014; WHO, 2014b; Davies and Davies, 2010). Furthermore, a recent report by the WHO has indicated an alarming global problem with regards to bacteria becoming resistant to the spectrum of commercially available antibiotics (WHO, 2014b). However, current global monitoring efforts are limited to tracking antibiotic consumption and antibiotic resistant bacteria isolated from clinical and public health laboratories (Grundmann *et al.*, 2011).

Antibiotic resistance determinants (ARD) include antibiotic resistance genes (ARGs) and mobile genetic elements (MGEs), conferring AR and is a mechanism for the transfer of ARGs in bacteria, respectively. These ARD have been reported across all types of environmental systems such as soil, water, sediments, human and animal faecal samples, wastewater sludge and wastewaters (Li *et al.*, 2015b; Czekalski *et al.*, 2014; Port *et al.*, 2014; Chen *et al.*, 2013). Previous studies have already suggested the anthropogenic sources of ARGs with aquatic

environments acting as major reservoirs (Marti *et al.*, 2014). These findings highlight that ARGs can spread from environmental reservoirs to human pathogens. Specifically, urban wastewater treatment plants have been implicated as being a hotspot for horizontal ARGs transfers between bacteria from difference origins due to the mixing of domestic, industrial, agricultural and clinical waste (Rizzo *et al.*, 2013). Furthermore, subsequent dissemination of poorly treated wastewater effluents into aquatic sources have been shown to spread residual antibiotics, antibiotic resistance bacteria (ARB) and ARGs into aquatic environments (Kümmerer, 2009; Lupo *et al.*, 2012), providing conditions for horizontal exchange of ARGs and adding further selection pressure for MRB. The overall effects of wastewater treatment plants discharges on environmental resistome is poorly understood and has potentially far reaching influence on intensifying the antibiotic resistance problem.

Therefore, to gain better understanding of the types of ARGs that are present in different environmental scenarios in conjunction with the MGEs such as integrons, transposons, NGS has provided the opportunity to circumvent the limitations of conventional culturing and molecular techniques. Recent studies have taken two different approaches, which have been sequence-based or function-based analysis. Sequence-based approaches usually isolate and extracts the total DNA directly from an environmental sample followed by shotgun sequencing. The NGS data obtained is then compared to a reference database containing known ARGs sequences, allowing for prediction of the resistance potential of the sample studied. On the other hand, functional-based approaches have involved the cloning of randomly shared total DNA fragments isolated and extracted from an environmental sample into an expression vector and transformed into a suitable host (typically laboratory *E. coli* strains). The *E. coli* library is then screened by selection of the transformants on media with the addition of selected

antibiotics. Individually, these two approaches allow for the identification of highly divergent genes from known ARGs and provide direct evidence of resistance phenotypes associated with the expressed genes, respectively (Pehrsson *et al.*, 2013). However, combined, studies have led to the discovery of novel ARGs in various microbial communities across diverse environments such as in soil (Torres-Cortés *et al.*, 2011; Allen *et al.*, 2009; Riesenfeld *et al.*, 2004), freshwaters (Bengtsson-Palme *et al.*, 2014; Diaz-Torres *et al.*, 2006), human oral and gut microbiomes (Devirgiliis *et al.*, 2014; Hu *et al*., 2013; Sommer *et al.*, 2009), animal gut microbiomes (Kazimierczak *et al.*, 2009) and wastewater treatment plants (Walsh, 2013; Uyaguari *et al.*, 2011; Parsley *et al.*, 2010; Mori *et al.*, 2008).

With further advances in the computational analysis, in terms of quality, accuracy and speed, NGS data should be expected to provide some novel and unparalleled insights from large scale studies assessing the threats posed by AR. This is further enhanced by the efforts of the scientific community in their contribution to the vast number of publicly available environmental microbiomes and current developments of bioinformatic tools for analysis such as the Comprehensive Antibiotic Resistance Database (McArthur *et al.*, 2013), the Antibiotic Resistance Database (Liu and Pop, 2009), the beta-lactamase database (Danishuddin *et al.*, 2013), ResFinder (Zankari *et al.*, 2012) and ARG-ANNOT (Gupta *et al.*, 2014) to name a few. Recently, Gibson *et al.* (2015) developed Resfams, a curated protein family database and associated profile hidden Markov models (HMMs) to determine the relationship between environmental and human associated resistome. The database is organized by ontology specifically applied to AR functions with a subset of these AR proteins functionally verified using protein assays. By using HMM and consensus models for functional annotation rather than the pairwise sequence alignment to AR databases, significant increases in prediction

sensitivity and specificity was reported. Gibson *et al.* (2015) investigated the AR profiles and occurrence patterns of microbial communities in aquatic environments through the use of network analysis. The study was able to determine the ARGs profiles as well as the co-occurrence patterns in 50 samples, spanning human, faecal, water, soil, sediments, wastewater and sludge samples. Aminoglycoside, bacitracin, beta-lactam, chloramphenicol, macrolide-lincosamide-streptogramin, quinolone, sulphonamide and tetracycline resistance genes were observed in high abundance and were associated with antibiotics commonly administered in human or veterinary medicine. Furthermore, abundances of these ARGs were up to three-fold higher in the most heavily anthropogenic-impacted environments. Similarly, a study by (Chen *et al.*, 2013) revealed higher diversity of ARGs for sulphonamides, fluoroquinolones, and aminoglycosides compared to pristine deep ocean samples from the South China Sea. Additionally, the study was able to show correlations between the distribution of ARGs and MGEs, where MGEs function as vectors for the dissemination of ARGS in the aquatic environment. A recent study suggested the use of a multivariate index as a means for quantification of the AR potential in publically available metagenomic data (Port *et al.*, 2014). The index is based on the abundance of ARGs, MGEs, pathogenic potential and metal resistance genes, which are implicated in co-selection of ARGs. In line with studies examining the distribution of ARGs, the index was reported to be able to differentiate between natural aquatic and anthropogenic impacted environments. Hence, this study illustrates the utility of NGS for advanced characterization of AR in water quality monitoring and could assist in risk assessments (Port *et al.*, 2014).

While clinical settings are known to be the source of highest antibiotic and ARB loads, the natural environment has recently drawn attention as a reservoir of transferable ARGs

potentially implicating environments highly contaminated with ARGs as a human health risk (Ashbolt *et al.*, 2013). ARDs in particular are increasingly being viewed as pollutants and emphasis of their occurrence and distribution in natural environments are being considered in development of antibiotic surveillance frameworks (Berendonk *et al.*, 2015; FAO, 2011). Incorporating the NGS approach into frameworks for monitoring the threat of AR in aquatic environments provides a novel approach for environmental health monitoring and pushes boundaries on improving current risk assessment models.

1.4 Scope of the present study

Increased rates of urbanization, industrialization and population growth have contributed to the decrease of freshwater quality while global freshwater scarcity is already a crisis. South Africa is a water stressed country receiving an average rainfall of 450 mm per year, leading to reduced levels of runoff and availability of surface water (Department of Water Affairs, 2010). This has led to the utilization of alternative freshwater sources for daily anthropogenic activities. However, stresses such as salinity, eutrophication as well as sediment and acid mine run-off with additional anthropogenic contributions in biological and chemical pollution of these water sources impede their usage (Department of Environmental Affairs, 2016). This is further compounded by failure of a large number of existing wastewater treatment plants in meeting the national water quality standards leading to the discharge of poorly treated or untreated wastewater effluents, introducing excessive nutrients, phosphates and coliforms into the receiving water bodies in addition to other toxic pollutants from agricultural practices which uses pesticides, herbicides and fertilizers (Department of Environmental Affairs, 2016).

Furthermore, previous findings have already suggested wastewater treatment plants as major reservoirs for antibiotic resistance and as well as for dissemination of residual antibiotics (AB), antibiotic resistance genes (ARGs) and antibiotic resistant bacteria (ARB) into the environment (Rizzo *et al.*, 2013) which could pose public health risks. Antibiotics have played a crucial role in the clinical management of most bacterial diseases since the 1940s. However, these drugs have been short lived and gradually rendered obsolete with more cases of multidrug resistance organisms reported from the clinical sector. This would mean that the effectiveness of standard treatment protocols currently adopted in clinical and animal sectors for treatment of diseases causing pathogens is steadily declining and could lead to a projected increase in mortality rates and risk of infection to others (WHO, 2014b). In order to determine the point of concern within the wastewater treatment process, relative occurrence of ARGs and ARB need to be critically assessed in every treatment stage.

Current microbiological tools used to assess the spread of ARGs and enrichment of ARB in water have largely been based on traditional culture-dependent and several molecular-dependent methods, each with its own advantages and limitations (Rizzo *et al.*, 2013). However, the use of culture-dependent methods to comprehensively characterize and evaluate ARGs and ARB in complex microbial communities remains a challenge as it is a time consuming task with only less than 1% of environmental organisms culturable with current microbiological methods (Schmieder and Edwards, 2012). Advances in sequencing as well as computational developments over the past decade has allowed for cost effective high-throughput sequencing combined with metagenomic analysis of a given environmental sample with unparalleled insights. Therefore, this approach has been applied for the assessment of various environmental scenarios such as soil (Howe *et al.*, 2014) and river microbiota, communities in sediments from

an oil spill (Mason *et al.*, 2014), faecal microbiota (Ilmberger *et al.*, 2014), microbial diversity in aquatic settings (Doxey *et al.*, 2014) as well as clinical settings such as the human gut profiling project (Qin *et al.*, 2010), sputum analysis of patients with cystic fibrosis (Lim *et al.*, 2014), viral pathogens detection (Yang, 2014) and many others. In determining the complex genetic composition and diversity of wastewater samples with a relatively unbiased view as well as gaining understandings of the resistome of the communities and potential pathogens present, use of next-generation sequencing combined with metagenomic insight is a favourable approach. Therefore, this study investigated the role of two geographically varied full-scale wastewater treatment plants located in Durban as potential hotspots for antibiotic resistant bacteria and antibiotic resistance genes through a metagenomic approach.

1.4.1 Hypotheses

1.4.1.1 It was hypothesized that a metagenomic approach in the analysis of wastewater samples collected from two full-scale wastewater treatment plants in Durban, KwaZulu-Natal, South Africa will allow for taxonomic classification of the bacterial community present and functional genes characterized, as well as the determination of the frequency of the genes which play a role in antibiotic resistance.

1.4.1.2 It was further hypothesized that a shotgun metagenomic approach will allow for a comprehensive view of the microbial communities in wastewater and aquatic samples, thereby allowing for proper insights into the effectiveness of the treatment process carried out by the full-scale wastewater treatment plants and its

potential contributions to the dissemination of antibiotic resistant bacteria and antibiotic resistance genes into the environment.

1.4.2 Objectives

The following objectives were set:

1.4.2.1 To determine and analyse the microbial communities of wastewater and river samples using next-generation sequencing and bioinformatic tools.

1.4.2.2 To profile and analyse the microbial communities of wastewater and river samples using T-RFLP and appropriate statistical analysis tools.

1.4.2.3 To reconstruct draft genomes of organisms found in the metagenomes of the wastewater and river samples.

1.4.3 Aims

The following aims were pursued:

1.4.3.1 Taxonomically and phylogenetic classification of the microbial communities of the wastewater and river samples.

1.4.3.2 Functional characterization of the microbial communities of the wastewater and river samples.

1.4.3.3 Determination of the putative metabolic potential of the microbial communities by mapping enzymes annotated.

1.4.3.4 Determination of the various antibiotic resistance genes present in the microbial community.

1.4.3.5 Comparative analysis of the microbial community structure and function at various treatment stages.

1.4.3.6 Isolation and assemblage of draft genomes found in the metagenomic datasets of wastewater and river samples.

1.5 Present investigations

In order to achieve the stated objectives and aims, the present study was divided into seven chapters as described below:

Chapter 1: This chapter provides an overview of the global freshwater crisis as well as the global initiatives established by international organisations and government to safeguard these essential resources. In addition, in depth discussion on next-generation sequencing technology in terms of experimental design and analysis methods were highlighted. Furthermore, an overview of the application of NGS in microbial quantification, faecal and pathogen detections, chemical pollution and remediation strategies prediction, and detection of antibiotic resistance determinants were also provided.

Chapter 2: This chapter evaluated the efficiency of two independent full-scale wastewater treatment plants through comparative metagenomic analysis of the wastewater influent received and the treated wastewater effluents. Specifically, the microbial community, potential bacterial pathogens and antibiotic resistance genes were determined and compared.

Chapter 3: This chapter specifically focuses on the shifts in the microbial community composition, the functional potential and antibiotic resistome as a result of chlorination disinfection treatment at the two independent full-scale wastewater treatment plants through the use of a metagenomic approach.

Chapter 4: This chapter investigated the impact of the treated wastewater effluent discharges from two independent full-scale wastewater treatment plants on their respective effluent-receiving river bodies in the Durban area through the use of a metagenomic approach. Specifically, the changes of the antibiotic resistance genes, bacterial community structure and metabolic potential of the two rivers as impacted by treated effluent discharge was examined.

Chapter 5: This chapter investigated the bacterial community dynamics of two rivers receiving the treated wastewater effluent discharges from separate full-scale wastewater treatment plants in the Durban area though the use of terminal-restriction fragment length polymorphisms.

Chapter 6: This chapter demonstrated the usefulness and power of bioinformatics by applying the large amount of open-source bioinformatic tools to recover interesting draft genomes at the species level within all of the metagenomes analysed in this study.

Chapter 7: This chapter provides an overview of the significant findings reported within the various chapters. In addition, possible limitations and potential for future developments of the study are also highlighted.

1.6 References

Ahn, T., Chai, J., Pan, C., 2015. Sigma: strain-level inference of genomes from metagenomic analysis for biosurveillance. *Bioinformatics* 31, 170–177.

Albertsen, M., Hansen, L.B.S., Saunders, A.M., Nielsen, P.H., Nielsen, K.L., 2012. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J.* 6, 1094–1106.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–8.

Allen, H.K., Moe, L.A., Rodbumrer, J., Gaarder, A., Handelsman, J., 2009. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J.* 3, 243–251.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146.

Amos, G.C.A., Zhang, L., Hawkey, P.M., Gaze, W.H., Wellington, E.M., 2014. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. *Vet. Microbiol.* 171, 441–447.

Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonterghem, I., Hugenholtz, P., Tyson, G.W., 2014. CopyRighter: A rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2, 11.

Arjoon, A., Olaniran, A.O., Pillay, B., 2013. Enhanced 1,2-dichloroethane degradation in heavy metal co-contaminated wastewater undergoing biostimulation and bioaugmentation. *Chemosphere* 93, 1826–1834.

Arjoon, A., Olaniran, A.O., Pillay, B., 2015. Kinetics of heavy metal inhibition of 1,2-dichloroethane biodegradation in co-contaminated water. *J. Basic Microbiol.* 55, 277–284.

Ashbolt, N.J., Amézquita, A., Backhaus, T., Borriello, P., Brandt, K.K., Collignon, P., Coors, A., Finley, R., Gaze, W.H., Heberer, T., Lawrence, J.R., Larsson, D.G.J., McEwen, S.A., Ryan, J.J., Schönfeld, J., Silley, P., Snape, J.R., Van den Eede, C., Topp, E., 2013. Human Health Risk Assessment (HHRA) for environmental development and transfer of antibiotic resistance. *Environ. Health Perspect.* 121, 993–1001.

Ayomoh, M.K.O., Oke, S.A., Adedeji, W.O., Charles-Owaba, O.E., 2008. An approach to tackling the environmental and health impacts of municipal solid waste disposal in developing countries. *J. Environ. Manage.* 88, 108–114.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* 9, 75.

Barrell, R.A., Hunter, P.R., Nichols, G., 2000. Microbiological standards for water and their relationship to health risk. *Commun. Dis. Public Health* 3, 8–13.

Beesley, L., Moreno-Jiménez, E., Gomez-Eyles, J.L., Harris, E., Robinson, B., Sizmur, T., 2011. A review of biochars' potential role in the remediation, revegetation and restoration of contaminated soils. *Environ. Pollut.* 159, 3269–3282.

Bengtsson-Palme, J., Boulund, F., Fick, J., Kristiansson, E., Larsson, D.G.J., 2014. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Antimicrob. Resist. Chemother.* 5, 648.

Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., Bürgmann, H., Sørum, H., Norström, M., Pons, M.-N., Kreuzinger, N., Huovinen, P., Stefani, S., Schwartz, T., Kisand, V., Baquero, F., Martinez, J.L., 2015. Tackling antibiotic resistance: The environmental framework. *Nat. Rev. Microbiol*. 13, 310–317.

Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* 47, 1945–1951.

Blair, J.M.A., Webber, M.A., Baylay, A.J., Ogbolu, D.O., Piddock, L.J. V., 2014. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13, 42–51.

Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., Corbeil, J., 2012. Ray Meta: Scalable *de novo* metagenome assembly and profiling. *Genome Biol.* 13, R122.

Brady, A., Salzberg, S.L., 2009. Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* 6, 673–676.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60.

Bunn, S.E., 2016. Grand challenge for the future of freshwater ecosystems. *Front. Environ. Sci.* 4, 21.

Cabral, J.P.S., 2010. Water Microbiology. Bacterial Pathogens and Water. *Int. J. Environ. Res. Public Health* 7, 3657–3703.

Cai, L., Ju, F., Zhang, T., 2014. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl. Microbiol. Biotechnol.* 98, 3317–3326.

Cai, L., Zhang, T., 2013. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environ. Sci. Technol.* 47, 5433–5441.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

Carew, M.E., Pettigrove, V.J., Metzeling, L., Hoffmann, A.A., 2013. Environmental monitoring using next generation sequencing: Rapid identification of macroinvertebrate bioindicator species. *Front. Zool.* 10, 45.

Chatterji, S., Yamazaki, I., Bai, Z., Eisen, J.A., 2008. CompostBin: A DNA composition-based algorithm for binning environmental shotgun reads in research in computational molecular biology, Vingron, M. and Wong, L. (eds.), 1st edn., Lecture Notes in Computer Science. Springer, Berlin, Germany.

Chen, B., Yang, Y., Liang, X., Yu, K., Zhang, T., Li, X., 2013. Metagenomic profiles of antibiotic resistance genes (ARGs) between human impacted estuary and deep ocean sediments. *Environ. Sci. Technol.* 47, 12753–12760.

Cheng, L., Quek, C.Y.J., Sun, X., Bellingham, S.A., Hill, A.F., 2013. The detection of microRNA associated with Alzheimer's disease in biological fluids using next-generation sequencing technologies. *Front. Genet.* 4, 150.

Clements, A., Young, J.C., Constantinou, N., Frankel, G., 2012. Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes* 3, 71–87.

Coleman, M.L., Chisholm, S.W., 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18634–18639.

Consortium, H.M.P., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.

Czekalski, N., Gascón Díez, E., Bürgmann, H., 2014. Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. *ISME J.* 8, 1381–1390.

Danishuddin, M., Hassan Baig, M., Kaushal, L., Khan, A.U., 2013. BLAD: A comprehensive database of widely circulated beta-lactamases. *Bioinformatics* 29, 2515–2516.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A. V, Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., Turnbaugh, P.J., 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563.

Davies, J., Davies, D., 2010. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433.

DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M., Williams, C., Reich, M., Winckler, W., Getz, G., 2012. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530–1532.

Department of Environmental Affairs, 2016. 2nd South Africa Environment Outlook - a report on the state of the environment. Pretoria, South Africa.

Department of Water Affairs, 2010. Integrated Water Resource Planning for South Africa: A situation analysis. Pretoria, South Africa.

Devirgiliis, C., Zinno, P., Stirpe, M., Barile, S., Perozzi, G., 2014. Functional screening of antibiotic resistance genes from a representative metagenomic library of food fermenting microbiota. *Biomed Res. Int.* 2014, 9.

Diaz-Torres, M.L., Villedieu, A., Hunt, N., McNab, R., Spratt, D.A., Allan, E., Mullany, P., Wilson, M., 2006. Determining the antibiotic resistance potential of the indigenous oral microbiota of humans using a metagenomic approach. *FEMS Microbiol. Lett.* 258, 257–262.

Diaz, N.N., Krause, L., Goesmann, A., Niehaus, K., Nattkemper, T.W., 2009. TACOA: Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics* 10, 56.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., Banfield, J.F., 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10, R85.

Dolan, T., Howsam, P., Parsons, D.J., 2012. Diffuse pesticide pollution of drinking water sources: impact of legislation and UK responses. *Water Policy* 14, 680.

Doxey, A.C., Kurtz, D.A., Lynch, M.D.J., Sauder, L.A., Neufeld, J.D., 2014. Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. *ISME J.* 9, 461–471.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.-J., Richter, R.A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J.D., Haft, D.H., Halpern, A.L., Lasken, R.S., Nealson, K., Friedman, R., Venter, J.C., 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199.

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

Engel, A.S., Gupta, A.A., 2014. Regime shift in sandy beach microbial communities following Deepwater Horizon oil spill remediation efforts. *PLoS One* 9, e102934.

Esling, P., Lejzerowicz, F., Pawlowski, J., 2015. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* 43, 2513–2524.

Evangelista, K. V, Coburn, J., 2010. *Leptospira* as an emerging pathogen: A review of its biology, pathogenesis and host immune responses. *Future Microbiol.* 5, 1413–1425.

FAO, 2004. Good Agricultural Practices – a working concept. Food And Agriculture Organization of the United Nations, Rome, Italy.

FAO, 2011. Guidelines for risk analysis of foodborne antimicrobial resistance. URL: http://www.fao.org/food/food-safety-quality/a-z-index/antimicrobial/en/ (accessed 5.16.16).

Fogden, J., Wood, G., 2009. Access to safe drinking water and its impact on global economic growth, 1st edn. HaloSource, Inc., Washington, USA.

Francis, O.E., Bendall, M., Manimaran, S., Hong, C., Clement, N.L., Castro-Nallar, E., Snell, Q., Schaalje, G.B., Clement, M.J., Crandall, K.A., Johnson, W.E., 2013. Pathoscope: Species identification and strain attribution with unassembled sequencing data. *Genome Res.* 23, 1721–1729.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., Delong, E.F., 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3805–3810.

Gerlach, W., Stoye, J., 2011. Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Res.* 39, e91.

Gibson, M.K., Forsberg, K.J., Dantas, G., 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 9, 207–216.

Gieg, L.M., Fowler, S.J., Berdugo-Clavijo, C., 2014. Syntrophic biodegradation of hydrocarbon contaminants. *Curr. Opin. Biotechnol.* 27, 21–9.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., Moran, M.A., 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J.* 5, 461–472.

Gilbert, J.A., Dupont, C.L., 2011. Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* 3, 347–71.

Gomez-Alvarez, V., Revetta, R.P., Domingo, J.W.S., 2012. Metagenomic analyses of drinking water receiving different disinfection treatments. *Appl. Environ. Microbiol.* 78, 6095–6102.

Green, H.C., Haugland, R.A., Varma, M., Millen, H.T., Borchardt, M.A., Field, K.G., Walters, W.A., Knight, R., Sivaganesan, M., Kelty, C.A., Shanks, O.C., 2014. Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl. Environ. Microbiol.* 80, 3086–3094.

Grundmann, H., Klugman, K.P., Walsh, T., Ramon-Pardo, P., Sigauque, B., Khan, W., Laxminarayan, R., Heddini, A., Stelling, J., 2011. A framework for global surveillance of antibiotic resistance. *Drug Resist. Updat.* 14, 79–87.

Gumede, N., Lentsoane, O., Burns, C.C., Pallansch, M., Gourville, E. de, Yogolelo, R., Muyembe-Tamfum, J.J., Puren, A., Schoub, B.D., Venter, M., 2013. Emergence of vaccine-derived Polioviruses, Democratic Republic of Congo, 2004–2011. *Emerg. Infect. Dis.* 19. doi:10.3201/eid1910.130028

Guo, G., Wu, F., Xie, F., Zhang, R., 2012. Spatial distribution and pollution assessment of heavy metals in urban soils from southwest China. *J. Environ. Sci.* 24, 410–418.

Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L., Rolain, J.-M., 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* 58, 212–220.

Haas, C.N., Rose, J.B., Gerba, C.P., 2014. Quantitative microbial risk assessment, 2nd edn. John Wiley & Sons, Inc., New Jersey, C.A.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z.D., Wada, Y., Wisser, D., 2014. Global water resources affected by human interventions and climate change. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3251–3256.

Haider, B., Ahn, T.-H., Bushnell, B., Chai, J., Copeland, A., Pan, C., 2014. Omega: An overlap-graph *de novo* assembler for metagenomics. *Bioinformatics* 30, 2717–2722.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., Tyson, G.W., 2013. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 500, 567–570.

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl. Environ. Microbiol.* 80, 6771–6781.

Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S., Barry, K., Tringe, S.G., Watson, D.B., He, Z., Hazen, T.C., Tiedje, J.M., Rubin, E.M., Zhou, J., 2010. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* 4, 660–672.

Hoff, K.J., Lingner, T., Meinicke, P., Tech, M., 2009. Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res.* 37, W101–105.

Horan, N.J., 2003. Handbook of water and wastewater microbiology, 1st edn. Academic Press, London, U.K.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M., Brown, C.T., 2014. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl. Acad. Sci.* 111, 4904–4909.

Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., Pan, Y., Li, J., Zhu, L., Wang, X., Meng, Z., Zhao, F., Liu, D., Ma, J., Qin, N., Xiang, C., Xiao, Y., Li, L., Yang, H., Wang, J., Yang, R., Gao, G.F., Wang, J., Zhu, B., 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* 4, 2151.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., Pace, N., Stahl, D., Lane, D., Olsen, G., Handelsman, J., Rondon, M., Brady, S., Clardy, J., Goodman, R., Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J., Wheeler, D., Mitchell, A., Chang, H., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., Mackelprang, R., Waldrop, M., DeAngelis, K., David, M., Chavarria, K., Blazewicz, S., Kanehisa, M., Goto, S., Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., Jansson, J., Buchfink, B., Xie, C., Huson, D., Huson, D., Mitra, S., Weber, N., Ruscheweyh, H.,

Schuster, S., Huson, D., Auch, A., Qi, J., Schuster, S., Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., Overbeek, R., Olson, R., Pusch, G., Olsen, G., Davis, J., Disz, T., Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Willmann, M., El-Hadidi, M., Huson, D., Schütz, M., Weidenmaier, C., Autenrieth, I., Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., Wood, D., Salzberg, S., Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., Glass, E., Wilkening, J., Wilke, A., Antonopoulos, D., Meyer, F., Peabody, M., Rossum, T. Van, Lo, R., Brinkman, F., Lindgreen, S., Adair, K., Gardner, P., Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Servick, K., Buchfink, B., Huson, D., Xie, C., Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Goodrich, J., Rienzi, S., Poole, A., Koren, O., Walters, W., Caporaso, J., Bray, R., Curtis, J., Treangen, T., Koren, S., Sommer, D., Liu, B., Astrovskaya, I., Ondov, B., Howe, A., Jansson, J., Malfatti, S., Tringe, S., Tiedje, J., Brown, C., Wang, Q., Fish, J., Gilman, M., Sun, Y., Brown, C., Tiedje, J., 2016. MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput Biol* 12, 4–12.

Hyatt, D., Chen, G., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. B*MC Bioinformatics* 11, 119.

Ibekwe, A.M., Leddy, M., Murinda, S.E., 2013. Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. *PLoS One* 8, e79490.

Ilmberger, N., Güllert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., Alawi, M., Poehlein, A., Chow, J., Turaev, D., Rattei, T., Schmeisser, C., Salomon, J., Olsen, P.B., Daniel, R., Grundhoff, A., Borchert, M.S., Streit, W.R., 2014. A Comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. *PLoS One* 9, e106707.

Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., Tyson, G.W., 2014. GroopM: An automated tool for the recovery of population genomes from related metagenomes. *PeerJ* 2, e603.

Jägerskog, A., Jønch Clausen, T., 2012. Feeding a thirsty world – Challenges and opportunities for water and food secure future. Stockholm International Water Institute, Stockholm, Sweden.

Ju, F., Guo, F., Ye, L., Xia, Y., Zhang, T., 2014. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ. Microbiol. Rep.* 6, 80–89.

Ju, F., Zhang, T., 2014. Novel microbial populations in ambient and mesophilic biogas-producing and phenol-degrading consortia unraveled by high-throughput sequencing. *Microb. Ecol.* 68, 235–246.

Ju, F., Zhang, T., 2015. Experimental design and bioinformatics analysis for the application of metagenomics in environmental sciences and biotechnology. *Environ. Sci. Technol.* 49, 12628–40.

Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., Bäckhed, F., 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–103.

Kazimierczak, K.A., Scott, K.P., Kelly, D., Aminov, R.I., 2009. Tetracycline resistome of the organic pig gut. *Appl. Environ. Microbiol.* 75, 1717–1722.

Kelly, L., Ding, H., Huang, K.H., Osburne, M.S., Chisholm, S.W., 2013. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J.* 7, 1827–1841.

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., Neufeld, J.D., 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl. Environ. Microbiol.* 80, 5717–5722.

Kent, W.J., 2002. BLAT--the BLAST-like alignment tool. *Genome Res.* 12, 656–664.

Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J.A., Hugenholtz, P., van der Lelie, D., Meyer, F., Stevens, R., Bailey, M.J., Gordon, J.I., Kowalchuk, G.A., Gilbert, J.A., 2012. Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol.* 30, 513–520.

Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., Mardis, E.R., 2013. The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38.

Koenig, J., Lee, M., Manefield, M., 2014. Aliphatic organochlorine degradation in subsurface environments. *Rev. Environ. Sci. Biotechnology* 14, 49–71.

Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., Turner, D.J., 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* 6, 291–295.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120.

Kraljevic, A., Geiger, M., 2012. Water conflict - Myth or reality? Berlin, Germany.

Kristiansen, R., Nguyen, H.T.T., Saunders, A.M., Nielsen, J.L., Wimmer, R., Le, V.Q., McIlroy, S.J., Petrovski, S., Seviour, R.J., Calteau, A., Nielsen, K.L., Nielsen, P.H., 2013. A metabolic model for members of the genus Tetrasphaera involved in enhanced biological phosphorus removal. *ISME J.* 7, 543–554.

Kughur, P.G., 2012. The effects of herbicides on crop production and environment in Makurdi Local Government area of Benu State, Nigeria. *J. Sustain. Dev. Africa* 4, 206–216.

Kumar, N. V, Menon, T., Pathipati, P., Cherian, K.M., 2013. 16S rRNA sequencing as a diagnostic tool in the identification of culture-negative endocarditis in surgically treated patients. *J. Heart Valve Dis.* 22, 846–849.

Kümmerer, K., 2009. Antibiotics in the aquatic environment – A review – Part II. *Chemosphere* 75, 435–441.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., Hugenholtz, P., 2008. A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* 72, 557–578.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Leclerc, H., Schwartzbrod, L., Dei-Cas, E., 2002. Microbial agents associated with waterborne diseases. *Crit. Rev. Microbiol.* 28, 371–409.

Li, A., Li, L., Zhang, T., 2015a. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. *Front. Microbiol.* 6, 1025.

Li, B., Ju, F., Cai, L., Zhang, T., 2015b. Profile and Fate of Bacterial Pathogens in Sewage Treatment Plants Revealed by High-Throughput Metagenomic Approach. *Environ. Sci. Technol.* 49, 10492–10502.

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., Zhang, T., 2015c. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* 9, 2490–2502.

Li, D., Liu, C., Luo, R., Sadakane, K., Lam, T., 2015d. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–6.

Li, H., Durbin R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Lim, Y.W., Evangelista, J.S., Schmieder, R., Bailey, B., Haynes, M., Furlan, M., Maughan, H., Edwards, R., Rohwer, F., Conrad, D., 2014. Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. *J. Clin. Microbiol.* 52, 425–437.

Liu, B., Gibbons, T., Ghodsi, M., Pop, M., 2010. MetaPhyler: Taxonomic profiling for metagenomic sequences, *in:* 2010 IEEE International conference on bioinformatics and biomedicine (BIBM). IEEE, pp. 95–100.

Liu, B., Pop, M., 2009. ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res.* 37, D443–D447.

Liu, R., Zhang, P., Wang, X., Chen, Y., Shen, Z., 2013. Assessment of effects of best management practices on agricultural non-point source pollution in Xiangxi River watershed. *Agric. Water Manag.* 117, 9–18.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., de Vargas, C., Lima-Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels-Lewis, S., Karsenti, E., Bork, P., Acinas, S.G., 2013. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* 16, 2659–2671.

Lu, X., Zhang, X.-X., Wang, Z., Huang, K., Wang, Y., Liang, W., Tan, Y., Liu, B., Tang, J., 2015. Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. *PLoS One* 10, e0125549.

Lupo, A., Coyne, S., Berendonk, T.U., 2012. Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. *Front. Microbiol.* 3, 18.

Ma, J., Prince, A., Aagaard, K.M., 2014. Use of whole genome shotgun metagenomics: a practical guide for the microbiome-minded physician scientist. *Semin. Reprod. Med.* 32, 5–13.

Magoč, T., Salzberg, S.L., 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.

Major, D.W., McMaster, M.L., Cox, E.E., Edwards, E.A., Dworatzek, S.M., Hendrickson, E.R., Starr, M.G., Payne, J.A., Buonamici, L.W., 2002. Field demonstration of successful bioaugmentation to achieve dechlorination of tetrachloroethene to ethene. *Environ. Sci. Technol.* 36, 5106–5116.

Mande, S.S., Mohammed, M.H., Ghosh, T.S., 2012. Classification of metagenomic sequences: methods and challenges. *Brief. Bioinform.* 13, 669–81.

Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W., Wommack, K.E., 2014. Caught in the middle with multiple displacement amplification: The myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* 2, 3.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., Ratner, A., Jacob, B., Pati, A., Huntemann, M., Liolios, K., Pagani, I., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG/M: The integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* 40, D123–129.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Pagani, I., Tringe, S., Huntemann, M., Billis, K., Varghese, N., Tennessen, K., Mavromatis, K., Pati, A., Ivanova, N.N., Kyrpides, N.C., 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, D568–5673.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.-M.A., Chu, K., Kyrpides, N.C., 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271–2278.

Marshall, C.W., Ross, D.E., Fichot, E.B., Norman, R.S., May, H.D., 2012. Electrosynthesis of commodity chemicals by an autotrophic microbial community. *Appl. Environ. Microbiol.* 78, 8412–8420.

Marti, E., Variatza, E., Balcazar, J.L., 2014. The role of aquatic ecosystems as reservoirs of antibiotic resistance. *Trends Microbiol.* 22, 36–41.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13, 31.

Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S.G., Dubinsky, E.A., Fortney, J.L., Han, J., Holman, H.-Y.N., Hultman, J., Lamendella, R., Mackelprang, R., Malfatti, S., Tom, L.M., Tringe, S.G., Woyke, T., Zhou, J., Rubin, E.M., Jansson, J.K., 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* 6, 1715–1727.

Masrevaniah, A., 2010. Prediction of eutrophication risk caused by N and P containing agriculture waste. *Int. J. Acad. Res.* 2, 191–193.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J. V, Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., Wright, G.D., 2013. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–57.

McHardy, A.C., Martín, H.G., Tsirigos, A., Hugenholtz, P., Rigoutsos, I., 2007. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* 4, 63–72.

McIlroy, S.J., Kristiansen, R., Albertsen, M., Karst, S.M., Rossetti, S., Nielsen, J.L., Tandoi, V., Seviour, R.J., Nielsen, P.H., 2013. Metabolic model for the filamentous "*Candidatus Microthrix parvicella*" based on genomic and metagenomic analyses. *ISME J.* 7, 1161–72.

McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N., Sogin, M.L., 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* 12, 378–392.

Mendes, M.P., Ribeiro, L., Nascimento, J., Condesso de Melo, T., Stigter, T.Y., Buxo, A., 2012. A groundwater perspective on the river basin management plan for central Portugal – developing a methodology to assess the potential impact of N fertilizers on groundwater bodies. *Water Sci. Technol.* 66, 2162.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386.

Mitra, S., Förster-Fromme, K., Damms-Machado, A., Scheurenbrand, T., Biskup, S., Huson, D.H., Bischoff, S.C., 2013. Analysis of the intestinal microbiota using SOLiD 16S rRNA gene sequencing and SOLiD shotgun sequencing. *BMC Genomics* 14, S16.

Mori, T., Mizuta, S., Suenaga, H., Miyazaki, K., 2008. Metagenomic screening for bleomycin resistance genes. *Appl. Environ. Microbiol.* 74, 6803–6805.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J., Gerstein, M., 2016. The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol.* 17, 53.

Muller, E.E.L., Glaab, E., May, P., Vlassis, N., Wilmes, P., 2013. Condensing the omics fog of microbial communities. *Trends Microbiol.* 21, 325–333.

Murray, D.C., Coghlan, M.L., Bunce, M., 2015. From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS One* 10, e0124671.

Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y., 2012. MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40, e155.

Nataro, J.P., Kaper, J.B., 1998. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.

Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L., Graf, J., 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One* 9, e94249.

Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M., Sogin, M.L., 2015. Sewage reflects the microbiomes of human populations. *MBio* 6, e02574.

Noguchi, H., Taniguchi, T., Itoh, T., 2008. MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15, 387–396.

Okeke, I.N., 2009. Diarrheagenic *Escherichia coli* in sub-Saharan Africa: Status, uncertainties and necessities. *J. Infect. Dev. Ctries.* 3, 817–842.

Olaniran, A.O., Igbinosa, E.O., 2011. Chlorophenols and other related derivatives of environmental concern: Properties, distribution and microbial degradation processes. *Chemosphere* 83, 1297–1306.

Ong, S.H., Kukkillaya, V.U., Wilm, A., Lay, C., Ho, E.X.P., Low, L., Hibberd, M.L., Nagarajan, N., 2013. Species identification and profiling of complex microbial communities using shotgun Illumina sequencing of 16S rRNA amplicon sequences. *PLoS One* 8, e60811.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–214.

Pace, N.R., 2009. Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* 73, 565–576.

Pandey, P.K., Kass, P.H., Soupir, M.L., Biswas, S., Singh, V.P., 2014. Contamination of water resources by pathogenic bacteria. *A.M.B. Express* 4, 51.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ Prepr.* 2, e554v1

Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper, W.F., Liles, M.R., 2010. Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl. Environ. Microbiol.* 76, 3753–3757.

Patel, R.K., Jain, M., 2012. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One* 7, e30619.

Payment, P., Locas, A., 2011. Pathogens in water: Value and limits of correlation with microbial indicators. *Ground Water* 49, 4–11.

Pehrsson, E.C., Forsberg, K.J., Gibson, M.K., Ahmadi, S., Dantas, G., 2013. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* 4, 145.

Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428.

Port, J.A., Cullen, A.C., Wallace, J.C., Smith, M.N., Faustman, E.M., 2014. Metagenomic frameworks for monitoring antibiotic resistance in aquatic environments. *Environ. Health Perspect.* 122, 222–228.

Prakash, T., Taylor, T.D., 2012. Functional assignment of metagenomic data:Challenges and applications. *Brief. Bioinform.* 13, 711–727.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K. Le, Leclerc, M., Maguin,

E., Minardi, R.M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.

Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P., Oyola, S.O., 2012. Optimal enzymes for amplifying sequencing libraries. *Nat. Methods* 9, 10–11.

Raes, J., Korbel, J.O., Lercher, M.J., von Mering, C., Bork, P., 2007. Prediction of effective genome size in metagenomic samples. *Genome Biol.* 8, R10.

Rho, M., Tang, H., Ye, Y., 2010. FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191.

Ribbe, L., Delgado, P., Salgado, E., Flügel, W.-A., 2008. Nitrate pollution of surface water induced by agricultural non-point pollution in the Pocochay watershed, Chile *in:* 10th IWA International specialized conference on diffuse pollution and sustainable basin management 18–22 September 2006, Istanbul, Turkey 226, 13–20.

Riesenfeld, C.S., Goodman, R.M., Handelsman, J., 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* 6, 981–989.

Rivers, A.R., Sharma, S., Tringe, S.G., Martin, J., Joye, S.B., Moran, M.A., 2013. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *ISME J.* 7, 2315–2329.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I., Fatta-Kassinos, D., 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Sci. Total Environ.* 447, 345–360.

Rockström, J., Lannerstad, M., Falkenmark, M., 2007. Assessing the water challenge of a new green revolution in developing countries. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6253–6260.

Rosseel, T., Van Borm, S., Vandenbussche, F., Hoffmann, B., van den Berg, T., Beer, M., Höper, D., 2013. The origin of biased sequence depth in sequence-independent nucleic acid amplification and optimization for efficient massive parallel sequencing. *PLoS One* 8, e76144.

Satinsky, B.M., Crump, B.C., Smith, C.B., Sharma, S., Zielinski, B.L., Doherty, M., Meng, J., Sun, S., Medeiros, P.M., Paul, J.H., Coles, V.J., Yager, P.L., Moran, M.A., 2014. Microspatial gene expression patterns in the Amazon River Plume. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11085–11090.

Satinsky, B.M., Fortunato, C.S., Doherty, M., Smith, C.B., Sharma, S., Ward, N.D., Krusche, A. V, Yager, P.L., Richey, J.E., Moran, M.A., Crump, B.C., 2015. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome* 3, 39.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., Quince, C., 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 43, e37.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.

Schmidt, T.S.B., Matias Rodrigues, J.F., von Mering, C., 2015. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ. Microbiol.* 17, 1689–1706.

Schmieder, R., Edwards, R., 2011a. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864.

Schmieder, R., Edwards, R., 2011b. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6, e17288.

Schmieder, R., Edwards, R., 2012. Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol.* 7, 73–89.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814.

Sergeant, M.J., Constantinidou, C., Cogan, T., Penn, C.W., Pallen, M.J., 2012. High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. *PLoS One* 7, e38094.

Shanafield, M., Rosen, M., Saito, L., Chandra, S., Lamers, J., Nishonov, B., 2010. Identification of nitrogen sources to four small lakes in the agricultural region of Khorezm, Uzbekistan. *Biogeochemistry* 101, 357–368.

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl. Environ. Microbiol.* 79, 2906–2913.

Shanks, O.C., White, K., Kelty, C.A., Hayes, S., Sivaganesan, M., Jenkins, M., Varma, M., Haugland, R.A., 2010. Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 76, 1359–1366.

Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., Banfield, J.F., 2012. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res.* 23, 111–120.

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5, 209.

Sheik, C.S., Jain, S., Dick, G.J., 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ. Microbiol.* 16, 304–317.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., Novod, N., Hung, D.T., Gnirke, A., Garber, M., Guttman, M., Livny, J., 2015. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* 12, 323–325.

Simpson, J.M., Santo Domingo, J.W., Reasoner, D.J., 2002. Microbial source tracking: State of the science. *Environ. Sci. Technol.* 36, 5279–5288.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132.

Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: Evaluating the illumina platform. *PLoS One* 10, e0116955.

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., Newton, K., Brown, M.H., Mitchell, J.G., 2012. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. *Environ. Microbiol.* 14, 240–253.

Sommer, M.O.A., Dantas, G., Church, G.M., 2009. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325, 1128–1131.

Stefanowicz, A.M., Niklińska, M., Laskowski, R., 2009. Pollution-induced tolerance of soil bacterial communities in meadow and forest ecosystems polluted with heavy metals. *Eur. J. Soil Biol.* 45, 363–369.

Strous, M., Kraft, B., Bisdorf, R., Tegetmeyer, H.E., 2012. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* 3, 410.

Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P., Yasunaga, T., Kamagata, Y., Liu, W., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ. Microbiol.* 14, 441–452.

Tan, B., Fowler, S.J., Abu Laban, N., Dong, X., Sensen, C.W., Foght, J., Gieg, L.M., 2015a. Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. *ISME J.* 9, 2028–2045.

Tan, B., Ng, C., Nshimyimana, J.P., Loh, L.L., Gin, K.Y.H., Thompson, J.R., 2015b. Next-generation sequencing (NGS) for assessment of microbial water quality: Current progress, challenges, and future opportunities. *Front. Microbiol.* 6, 1027.

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., Glöckner, F.O., 2004. TETRA: A web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics* 5, 163.

Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - A guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 3.

Torres-Cortés, G., Millán, V., Ramírez-Saad, H.C., Nisa-Martínez, R., Toro, N., Martínez-Abarca, F., 2011. Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environ. Microbiol.* 13, 1101–1114.

Torres-García, W., Zheng, S., Sivachenko, A., Vegesna, R., Wang, Q., Yao, R., Berger, M.F., Weinstein, J.N., Getz, G., Verhaak, R.G.W., 2014. PRADA: Pipeline for RNA sequencing data analysis. *Bioinformatics* 30, 2224–2226.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.

Ulloa, O., Canfield, D.E., DeLong, E.F., Letelier, R.M., Stewart, F.J., 2012. Microbial oceanography of anoxic oxygen minimum zones. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15996–16003.

UN, 2015a. The Millennium Development Goals Report. New York, NY.

UN, 2015b. Water for a sustainable world.

UNEP, 2005. Water and Wastewater reuse: An Environmentally Sound Approach for Sustainable Urban Water Management.

UNEP, 2008. Vital Water Graphics - An Overview of the State of the World's Fresh and Marine Waters. Nairobi, Kenya.

Unno, T., Jang, J., Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.-S., Chun, J., Hur, H., 2010. Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environ. Sci. Technol.* 44, 7777–7782.

USEPA, 2009. Review of published studies to characterize relative risks from different sources of fecal contamination in recreational waters. Washington, USA.

USEPA, 2013. Method 1609: *Enterococci* in water by TaqMan® quantitative polymerase chain reaction (qPCR) with internal amplification control (IAC) assay. Washington, USA.

Uyaguari, M.I., Fichot, E.B., Scott, G.I., Norman, R.S., 2011. Characterization and quantitation of a novel β-Lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. *Appl. Environ. Microbiol.* 77, 8226–8233.

Valverde, J.R., Mellado, R.P., 2013. Analysis of metagenomic data containing high biodiversity levels. *PLoS One* 8, e58118.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014a. Ten years of next-generation sequencing technology. Trends Genet. 30, 418–426. doi:10.1016/j.tig.2014.07.001

van Dijk, E.L., Jaszczyszyn, Y., Thermes, C., 2014b. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp. Cell Res.* 322, 12–20.

Vandewalle, J.L., Goetz, G.W., Huse, S.M., Morrison, H.G., Sogin, M.L., Hoffmann, R.G., Yan, K., McLellan, S.L., 2012. *Acinetobacter*, *Aeromonas* and *Trichococcus* populations dominate the microbial community within urban sewer infrastructure. *Environ. Microbiol.* 14, 2538–2552.

Varela, A.R., Manaia, C.M., 2013. Human health implications of clinically relevant bacteria in wastewater habitats. *Environ. Sci. Pollut. Res. Int.* 20, 3550–3569.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561.

Walsh, F., 2013. Investigating antibiotic resistance in non-clinical environments. *Front. Microbiol.* 4, 19.

Wang, H., Du, P., Li, J., Zhang, Y., Zhang, W., Han, N., Woo, P.C.Y., Chen, C., 2014. Comparative analysis of microbiome between accurately identified 16S rDNA and quantified bacteria in simulated samples. *J. Med. Microbiol.* 63, 433–440.

Wang, L., Wang, S., Li, W., 2012. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185.

Wang, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012a. MetaCluster 4.0: A novel binning algorithm for NGS reads and huge number of species. *J. Comput. Biol.* 19, 241–9.

Wang, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012b. MetaCluster 5.0: A two-round binning approach for metagenomic data for low-abundance species in a noisy sample. *Bioinformatics* 28, i356–i362.

Wang, Y., Sheng, H., He, Y., Wu, J., Jiang, Y., Tam, N.F., Zhou, H., 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* 78, 8264–8271.

WHO, 2001. Water Quality: Guidelines, Standards and Health, First. ed. IWA Publishing, London, UK.

WHO, 2011. Guidelines for drinking-water quality, 4th edn., Geneva, Switzerland.

WHO, 2014a. Burden of disease and cost-effectiveness estimates. URL: http://www.who.int/water_sanitation_health/diseases/burden/en/ (accessed 5.10.16).

WHO, 2014b. Antimicrobial resistance: global report on surveillance 2014. URL: http://www.who.int/drugresistance/documents/surveillancereport/en/ (accessed 4.28.16).

WHO/IWA, 2011. Valuing water, valuing livelihoods: Guidance on social cost-benefit analysis of drinking-water interventions, with special reference to small community supplies. WHO/IWA, Geneva, Switzerland.

WHO/UNICEF, 2012. Progress on drinking water and sanitation: 2012 Update.

Wooley, J., 2007. The new science of metagenomics: Revealing the secrets of our microbial planet, 1st edn. National Academies Press, Washington, USA.

World Economic Forum, 2015. Global Risks 2015, 10th edn. Geneva, Switzerland.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., Long, P.E., Banfield, J.F., 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–5.

Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W., 2011. WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12, 444.

Wu, Y., Tang, Y., Tringe, S.G., Simmons, B.A., Singer, S.W., 2014. MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26.

Wylie, K.M., Truty, R.M., Sharpton, T.J., Mihindukulasuriya, K.A., Zhou, Y., Gao, H., Sodergren, E., Weinstock, G.M., Pollard, K.S., 2012. Novel bacterial taxa in the human microbiome. *PLoS One* 7, e35294.

Xia, Y., Wang, Y., Fang, H.H.P., Jin, T., Zhong, H., Zhang, T., 2014. Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci. Rep.* 4, 6708.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T., Li, Y., Xu, X., Wong, G.K., Wang, J., 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666.

Yampara-Iquise, H., Zheng, G., Jones, J.E., Carson, C.A., 2008. Use of a *Bacteroides thetaiotaomicron*-specific alpha-1-6, mannanase quantitative PCR to detect human faecal pollution in water. *J. Appl. Microbiol.* 105, 1686–1693.

Yang, J., 2014. Viral pathogens in clinical samples by use of a metagenomic approach, *in:* Nelson, K.E. (Ed.), Encyclopedia of Metagenomics. Springer, New York, U.S.A pp. 1–6.

Yang, Y., Li, B., Ju, F., Zhang, T., 2013. Exploring Variation of Antibiotic Resistance Genes in Activated Sludge over a Four-Year Period through a Metagenomic Approach. Environ. Sci. Technol. 47, 10197–10205. doi:10.1021/es4017365

Ye, L., Zhang, T., 2011. Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. *Environ. Sci. Technol.* 45, 7173–7179.

Ye, L., Zhang, T., 2013. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl. Microbiol. Biotechnol.* 97, 2681–2690.

Yergeau, E., Lawrence, J.R., Sanschagrin, S., Waiser, M.J., Korber, D.R., Greer, C.W., 2012. Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl. Environ. Microbiol.* 78, 7626–7637.

Yu, K., Zhang, T., 2012. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PLoS One* 7, e38183.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644.

Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.

Zhao, Y., Tang, H., Ye, Y., 2012. RAPSearch2: A fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 28, 125–126.

Zhou, Q., Su, X., Ning, K., 2014. Assessment of quality control approaches for metagenomic data analysis. *Sci. Rep.* 4, 6957.

Zhu, W., Lomsadze, A., Borodovsky, M., 2010. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res.* 38, e132.

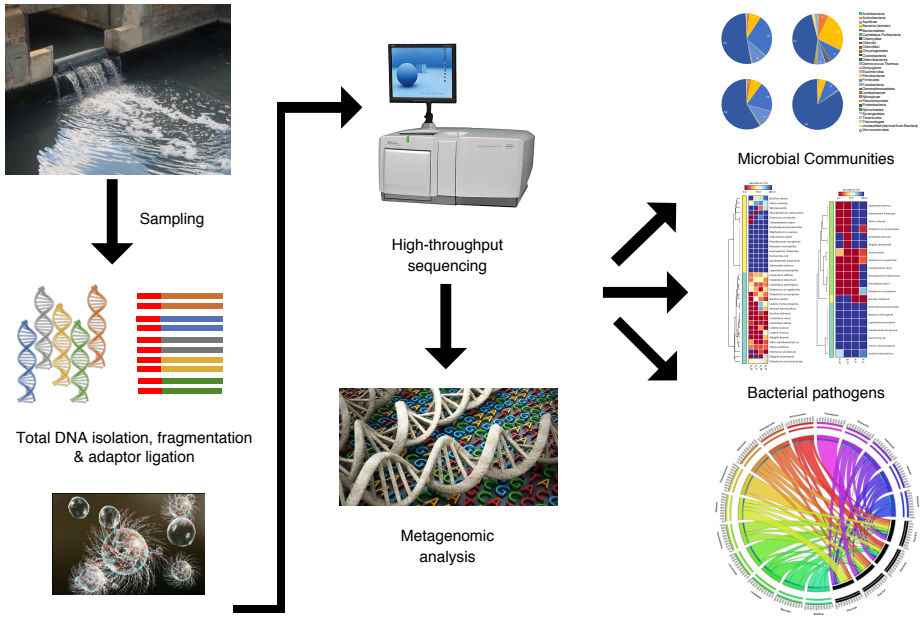
Chapter Two

Manuscript formatted and submitted to *Science of the Total Environment*

2 Fate of bacterial pathogens, community composition and antibiotic resistance genes in two full-scale wastewater treatment plants as revealed by metagenomic analysis

2.1 Abstract

Monitoring of treated effluent from urban wastewater treatment plants (WWTPs) is important to ascertain their treatment efficiency as well as to avoid circulation of pathogens in the environment. This study applied a molecular approach to assess the efficiency of two different urban full-scale WWTPs in treating the influent samples received. The pathogenic bacterial population, community composition and antibiotic resistance genes (ARGs) were comprehensively profiled in both the influent and treated effluent samples of both WWTPs using whole metagenome shotgun pyrosequencing combined with metagenomic analysis. Overall community analysis revealed changes in the bacterial community composition from influent to effluent samples of both WWTPs, with members of the Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes phyla dominating the communities. Though accounting for a small percentage of the communities, up to 19 genera of pathogenic bacteria were still detected in the treated effluent, with *Acinetobacter baumannii* and *Legionella pneumophila* observed to dominate the population, which deserves more concerns. Furthermore, diverse antibiotic resistance determinants accounting for majority of clinically relevant classes of antibiotics were detected across the samples, further highlighting urban WWTPs as sources of antibiotic resistance. Results from this study indicate that urban WWTPs may act as an important source of bacterial pathogens and ARGs and could contribute to the increasing global problem of multidrug resistant pathogens. Additionally, the use of high-throughput sequencing approach may be a feasible yet reliable method for the comprehensive assessment of these environmental sources.



(Graphical abstract)

Keywords: Antibiotic resistance; Bacterial pathogens; Metagenomics; Microbial community; Wastewater treatment plants

2.2 Introduction

The treatment of domestic, industrial and hospital wastewaters rely heavily on urban wastewater treatment plants (WWTPs) infrastructures. Countries with high population density typically require numerous WWTPs with varying treatment capacities. For example, the KwaZulu-Natal province of South Africa has 11 districts and is serviced by 140 WWTPs with a total design capacity of 1090.8 Ml/day (Department of Water and Sanitation, 2015). Complex networks microbial communities have been utilised in the past decades for the effective treatment of wastewater. Specifically, reduction of various physicochemical parameters and pathogen load are the treatment targets before effluent is discharged into the receiving water bodies, or in some cases further treated for re-use in agriculture and aquaculture sectors (WHO, 2006). Of major concern is that bacterial pathogens may endure the treatment process, proliferate under favourable conditions and ultimately disseminate into the environment (Okoh et al., 2007). Previous findings have already suggested the anthropogenic sources of antibiotic resistance genes (ARGs) (Marti et al., 2014) with urban WWTPs implicated as hotspots for horizontal ARGs transfers between different bacteria (Rizzo et al., 2013). Together with this is the increasing concern about the emergence and selection of multidrug resistant bacterial pathogens in such settings (WHO, 2014). Dissemination of these organisms in the environment could lead to an increase in mortality rates and risk of infection to others as the standard treatment protocol currently adopted in clinical and animal sectors would become ineffective (WHO, 2014).

Current microbiological tools used to assess bacterial pathogens and ARGs in WWTPs have either been culture- or molecular-dependent (Schwartz et al., 2003), and include colony count (Wen et al., 2009), PCR-based (Toze, 1999), real-time PCR (Kim et al., 2013) and microarray

(Aw and Rose, 2012; Cao et al., 2011). However, these techniques all present with its own advantages and limitations (Rizzo et al., 2013). Recently, the molecular approach for ARGs has been extended from their identification and characterization (Li et al., 2010; Luo et al., 2010; Storteboom et al., 2010) to genetic elements (Gaze et al., 2011; Johnning et al., 2013) found in cultured organisms. This allows for further understanding of the dynamics involved with capturing and transferring of ARGs and generation of resistant bacteria in microbial communities found in WWTPs. However, the use of culture-dependent methods to comprehensively characterize bacterial pathogens and ARGs in the complex microbial communities of environmental samples remains a challenge as it is a time consuming task. Furthermore as less than 1% of environmental organisms remain cultivable *in vitro* in laboratories with current technologies, investigations with this approach limits the determination of the complete spectrum of the communities and the pathogenic population in wastewater samples (Schmieder and Edwards, 2012).

In the last few years, high-throughput sequencing combined with metagenomic analysis of a given sample has been considered a promising approach for the assessment of microbial communities from complex samples (Kristiansson et al., 2011; Wang et al., 2013; Zhang et al., 2011). Several studies have already confirmed the feasibility of this technique for the assessment of diverse environmental scenarios, including determination of soil (Howe et al., 2014) and river (Amos et al., 2014) microbiota, communities found in activated sludge from WWTPs (Ju et al., 2014), communities in sediments from an oil spill (Mason et al., 2014), faecal microbiota (Ilmberger et al., 2014), microbial diversity in aquatic settings (Doxey et al., 2014) as well as clinical settings such as the human gut profiling project (Qin et al., 2010), sputum analysis of patients with cystic fibrosis (Lim et al., 2014), viral pathogens detection

(Yang, 2014) and many others. Hence, in determining the complex genetic composition and diversity present in aquatic samples with a relatively unbiased view (Gomez-Alvarez et al., 2009), a metagenomic insight is a favourable approach.

Although previous studies have investigated the microbial diversity of wastewater samples, information about the pathogenic bacterial population and the antibiotic resistance available in relation to the communities are lacking. Hence, this study determined the bacterial diversity, pathogenic bacterial population and associated ARGs in the influent and effluent samples of two different WWTPs in the city of Durban, South Africa using a metagenomic approach. Specifically, the occurrence, abundance and diversity of the bacterial communities, pathogenic bacterial population and ARGs were comprehensively examined. The results from this study may help to extend our knowledge on the contribution of urban WWTPs in the dissemination of ARGs and the bacterial pathogens commonly found in WWTPs. It further demonstrated the feasibility of metagenomic approach in ARGs and microbial pathogen, and its potential application in evaluating environmental health.

2.3 Materials and Methods

2.3.1 Descriptions of WWTPs and sample collection

In this study, influents (IF) and effluents (EF) wastewater samples were collected from two full-scale urban WWTPs in the city of Durban, South Africa, hereon designated as Plant A and Plant B. Plant A has a capacity of 70 megaliters/day with an operational capacity of 96% and

use the activated sludge and diffused air liquid technologies with gravity thickening, anaerobic digestion and belt press dewatering sludge technologies. The plant receives a mixture of nearby domestic and industrial wastewaters and discharges its final effluent into a nearby river in a suburban location. Plant B has a capacity of 25 megaliters/day with an operational capacity of 76% and use the activated sludge liquid technology with anaerobic digestion and belt press dewatering sludge technologies. The plant receives a different mixture of domestic, industrial, hospital wastewaters and discharges its final effluent into a nearby river in a heavily urbanized location (Department of Water and Sanitation, 2015). Four samples were collected in 5 L plastic bottles pre-sterilized with 70% (v/v) ethanol and rinsed with 4 L of the sample at the various sampling sites prior to collection. Upon collection, the samples were transported on ice back to the laboratories within 3 h and stored at 4°C prior to DNA extraction which took place within 24 h.

2.3.2 Ethics statement

No special permits were required for this study. Permission for the collection of all wastewater samples from both WWTPs were granted by the relevant authorities of the respective WWTPs (Durban, South Africa).

2.3.3 Total DNA extraction and shotgun pyrosequencing

Prior to total DNA extraction, samples which visibly contained particles which would hinder DNA extraction was initially filtered through Whatman® filter paper #114 (Sigma-Aldrich, USA) to remove big particles and to allow for isolation of bacterial communities. Total DNA

of the bacterial populations from the samples were extracted using PowerWater™ DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. This kit is specifically designed for isolating bacterial DNA from environmental water samples and includes inhibitor removal technology aimed at removing humic acid and other organic matter commonly found in environmental samples that can interfere with downstream analyses. The resulting purity and quantity of the DNA preparation was determined using the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, USA) before samples were sent to the National Genomics Platform (Technology Innovation Agency, South Africa) for sequencing.

High-throughput whole metagenome shotgun pyrosequencing was conducted by the use of 454 GS FLX System (Roche, USA) with the GS FLX Titanium Rapid Library Preparation Kit (Roche, USA). The samples were barcoded by the use of GS FLX Titanium Rapid Library MID Adaptors Kit (Roche, USA) in order to enable multiplex sequencing. Five hundred microgram of DNA was used for library construction. Briefly, DNA was nebulized using nitrogen gas and purified using MinElute PCR Purification Kit (QIAGEN, Germany) according to manufacturer's instructions. Fragment end repair and attachment of adaptors to the samples was carried out according to standard protocol (Roche, USA). Thereafter, the quality of the libraries was assessed by the use of 2000 Bioanalyzer (Agilent Technologies, USA) using the High Sensitivity DNA Analysis Kit (Agilent Technologies, USA). This was followed by the measurement of the relative fluorescence of the DNA libraries on the GloMax-Multi Microplate Multimode Reader (Promega, USA) using 6-fluorescein amidite (6-FAM) standards and 6-FAM-labeled Multiplex Identifiers (MIDs). The results were uploaded onto the Rapid Library Quantitation Calculator (Roche, USA) and rapid libraries were made with a

final concentration of 1×10^7 molecules/µl. Emulsion PCR was carried out with the use of GS FLX Titanium MV emPCR Kit (Lib-L) (Roche, USA) according to manufacturer's instructions. Subsequently, four DNA copies per bead were enriched and used for the main sequencing run with *Escherichia coli* beads as a positive control (Roche, USA).

2.3.4 Data availability

All individual sequence reads have been deposited at the NCBI Short Read Archive (SRA) under the accessions no. SRR3629046 and SRR3629048 for IF and EF of Plant A datasets, respectively, and SRR3629051 and SRR3629057 for IF and EF of Plant B datasets, respectively.

2.3.5 Bioinformatic analysis

Raw data files from sequencing were de-multiplexed by removing the barcoded sequence and any secondary adapter sequences using in-house scripts. For initial quality control, raw reads were evaluated by CLC Genomics Workbench v.7.5 (CLC Bio, QIAGEN, Germany) quality control pipeline. In total, approximately 800,409 raw reads with an average length of 611 bp corresponding to approximately 390 million bp were generated for this study. Specifically, raw sequences generated were 158,797 and 227,021 reads for IF and EF, respectively at Plant A, and 234,034 and 180,557 reads for IF and EF, respectively at Plant B. The quality of all reads were within the acceptable standards for the sequencing platform (see Table S2.1 in the supplemental material). Therefore, no reads were trimmed, filtered or discarded before implementation of the MG-RAST v.3.6 pipeline for further quality control, prediction and

annotation (Meyer et al., 2008; Wilke et al., 2015). Ambiguous base filtering was first implemented by removing sequences with >5 ambiguous base pairs (bp). This was followed by length filtering where sequences with a length of >2 standard deviations from the mean were removed. Filtering was applied using SolexaQA (Cox et al., 2010) implemented in the MG-RAST pipeline. Upon ambiguous base and length filtering, approximately 9% and 5% of the reads from IF and EF, respectively at Plant A, and 9% and 17% of IF and EF, respectively at Plant B, were excluded from further analysis. Remaining reads which passed the quality control were allowed for further analysis.

2.3.6 Taxonomic annotations and classifications

For taxonomic identification and classifications, an initial BLAT search against reduced RNA database (90% identity clustered version of SILVA database) was performed and the rRNA-similar reads were then clustered at 97% identity with the longest read as the cluster representative. Thereafter, BLAT search of the cluster representative was performed against the M5rna database (see Table S2.1 in the supplemental material). M5rna is a ribosomal database with integration of SILVA, Greengenes and RDP databases (Meyer et al., 2008; Wilke et al., 2015). Taxonomic abundance was analyzed using the Lowest Common Ancestor (LCA) algorithm used in MEGAN (Huson et al., 2007) and implemented in MG-RAST with a maximum e-value cut-off of 1×10^{-5}, minimum identity of 60% and a minimum alignment length of 15, which is measured in amino acids for protein databases and bp for RNA databases (Zheng et al., 2015). This algorithm assigns each read to the LCA from the set of matching taxa when BLASTx is applied. For example, if a given read had sequence similarity to 3 different families within the same order, the read is assigned at the order level rather than assigning the read to a specific family. Hence, the LCA algorithm has been reported to have lower rates of

false positive assignments than the best hit classification algorithm implemented in the MG-RAST pipeline. However, this would result in a higher number of unspecific assignments or no hits in some cases (Huson et al., 2007).

2.3.7 Virulence factors annotation and bacterial pathogen classifications

To determine the human bacterial pathogens in the metagenomes in order to establish the pathogenic population within the microbial communities, annotation with the human pathogenic bacteria virulence factor database (VFDB v.2016) was used (Chen et al., 2015, 2005). A total of 2581 proteins sequences were downloaded from the VFDB website (http://www.mgc.ac.cn/VFs/), which were then grouped at the species level using in-house scripts. Compared to other databases, VFDB is a up-to-date database of virulence factors from various bacterial pathogens derived only from experimentally demonstrated and published literature (Chen et al., 2015, 2005). BLASTx against the database was carried out using DIAMOND tool v.0.7.11 (Buchfink et al., 2014) with a e-value cut-off of 1×10^{-5}. Read with its best BLAST hit to the protein sequence from VFDB was further filtered with a sequence similarity of >80% over an alignment of ≥50 amino acids (Lu et al., 2015; Zhang et al., 2011).

2.3.8 Statistical analysis

Significant differences between the metagenomic samples were determined by the use of statistical analysis of metagenomic profiles (STAMP) v.2.1.3 software package (Parks and Beiko, 2010). Statistical significance of differences between two samples (q values) was calculated on the basis of two-sided Fisher's exact test with Benjamin-Hochberg's false

discovery rate (FDR). The confidence intervals were determined by Newcombe-Wilson's method. Reads designated as unclassified were removed from analyses and only results with a *q* value of <0.05 were considered significant in this study (Pacchioni et al., 2014). Statistical significance of differences between more than two samples was calculated using multiple group ANOVA with Tukey-Kramer post-hoc tests at 0.95, an effect size (Eta-squared) and multiple test correction using the Benjamini-Hochberg FDR procedure (White et al., 2015).

2.3.9 Antibiotic resistance genes annotation and classifications

To identify ARGs in the metagenomes in order to establish the resistance profiles of the communities, annotation with the Comprehensive Antibiotic Resistance Database (CARD) was done (McArthur et al., 2013). The CARD database was chosen over other ARGs databases, such as Antibiotic Resistance Genes Online (Scaria et al., 2005), the microbial database of protein toxins, virulence factors, and antibiotic resistance genes (MvirDB) (Zhou et al., 2007), and Antibiotic Resistance Genes Database (Liu and Pop, 2009), because these databases are neither exhaustive nor regularly updated (Gupta et al., 2014). BLASTx against the database was carried out using DIAMOND tool v.0.7.11 (Buchfink et al., 2014) with a e-value cut-off of 1×10^{-5}. A read with its best BLAST hit was deemed ARG-like if the hit had a sequence similarity of >90% over an alignment of ≥25 amino acids (Chao et al., 2013; Wang et al., 2013). Although such a high similarity threshold excluded some divergent ARGs from the analysis, we still used a more conservative strategy and thus only focus on those that are highly similar to the known ARGs. The classification of ARG-like sequences was performed using the structured database of CARD (McArthur et al., 2013) and in-house written scripts. Visualization of the distributions of ARGs classes and their abundances in the total annotated ARGs from the metagenomes was constructed and visualized using Circos v.0.69.2

(Krzywinski et al., 2009). Using the Paleontological Statistics (PAST) software v.3 (Hammer et al., 2001) Mann-Whitney test was implemented to compare whether ARGs abundances were significantly different among the various sampled environments (Hu et al., 2013; Li et al., 2015b). Various diversity indices (Simpson, Shannon, Buzas and Gibson, Brillouin, Menhinick, Margalef, Equitability, Fisher's alpha, Berger-Parker and Chao 1) and principal coordinate analysis (PCoA) based on the Bray-Curtis dissimilarity coefficients of the abundance of ARGs types were determined using the 'vegan' package (Shen and Fulthorpe, 2015) from the program R (R Development Core Team, 2007) with a bootstrap of 1000.

2.4 Results

2.4.1 Taxonomic analysis of microbial communities

Taxonomic assignment of both predicted rRNA genes sequences and predicted proteins in all metagenomic datasets were conducted based on all the available annotation source databases in MG-RAST.

Domain distributions in the four wastewater samples, showed the expected dominance of bacteria across all metagenomes, accounting for 99.40% and 93.32% at Plant A, and 99.42% and 99.44% at Plant B for IF and EF, respectively. The remaining sequences were assigned to Archaea, Eukaryota and Viruses, with members of the Amphibia class dominating the eukaryotic domain in both the IF samples. However, members of the Oligohymenophorea class were observed to dominate the eukaryotic domain in the EF sample of Plant A, whilst

potentially novel members of the Streptophyta class (determined as unclassified) were observed to dominate the eukaryotic domain in TE sample of Plant B. On the other hand, members of the Methanobacteria class were observed to dominate the archaeal domain in the IF and EF samples of both plants. The remaining sequences were distributed amongst the viral domains in very low abundance (see Table S2.2 in the supplemental material).

For a better understanding of the bacterial community structure in all the metagenomes, taxonomic affiliation at different levels was analysed. At the phyla taxonomic level, bacterial communities were diverse with representatives from 28 phyla and 1 novel bacterial phylum (determined as unclassified, however, under the Bacterial domain level) observed across all samples, with the exception of the EF sample in Plant B which had 27 phyla and 1 novel bacterial phylum. Proteobacteria was observed as the dominant phylum in all samples accounting for 52.18% and 43.68% of the sequences in IF and EF communities of Plant A, respectively, and 58.12% and 83.81% of the sequences in IF and EF communities of Plant B, respectively (Figure 2.1). This was followed by Bacteroidetes, Firmicutes and then Actinobacteria phyla in IF of both plants. However, the relative abundance of bacterial dominance in EF samples of both plants was observed to be Bacteroidetes, Actinobacteria and Firmicutes. Statistical analysis using STAMP indicated that 24 out of the 28 phyla and 14 out of the 27 phyla were significantly different (q<0.05) at this taxonomic level for Plant A and B samples, respectively (see Figure S2.1 in the supplemental material). Notably, significant increases (q<0.05) in Proteobacterial members and significant decreases (q<0.05) of members belonging to the Bacteroidetes and Firmicutes phyla were observed from the EF communities relative to their IF communities in both WWTPs. Surprisingly, strains belonging to the Actinobacteria and several other phyla were significantly increased (q<0.05) in the EF

communities relative to their IF communities of Plant A (see Figure S2.1 in the supplemental material).

At the class taxonomic level, the bacterial community was diverse with representatives of more than 43 classes and 11 potentially novel classes observed across all samples. Bacteroidia dominated the IF wastewater in Plant A, followed by Gamma-, Beta- and Epsilon-proteobacteria, which accounted for 15.76%, 15.60% and 13.02% of the assigned reads, respectively, whilst Clostridia, Flavobacteria, Actinobacteria, Negativicutes, Delta- and Alpha-proteobacteria accounted for >1% to <6% of the assigned reads in the sample. Conversely, Betaproteobacteria dominated the EF wastewater of Plant A, followed by Alpha-, Gamma-proteobacteria and Actinobacteria, which accounted for 6.61%, 6.57% and 6.49% of the assigned reads, respectively, whilst Delta-proteobacteria, Bacteroidia, Clostridia, Flavobacteria, Planctomycetacia and Bacilli accounted for >1% to <4% of the assigned reads in the sample. For Plant B, IF sample was dominated by Gammaproteobacteria, followed by Beta-proteobacteria, Bacteroidia, Epsilon-proteobacteria and Clostridia, accounting for 14.44%, 11.74%, 11.10% and 6.35% of the assigned reads, respectively, whilst Flavobacteria, Actinobacteria, Negativicutes, Bacilli, Delta- and Alpha-proteobacteria accounted for >1% to <6% of the assigned reads in the sample. Conversely, Gammaproteobacteria dominated the EF sample, followed by Beta-proteobacteria, Flavobacteria and Alpha-proteobacteria, which accounted for 9.16%, 6.07% and 1.63% of the assigned reads, respectively (Figure 2.2). Although Bacteroidia dominated the community in the IF sample of Plant A, a significant (q<0.05) reduction was observed in the EF sample. Conversely, Beta-, Alpha-, Delta-proteobacteria, Actinobacteria and several other phyla were significantly (q<0.05) increased in the EF communities (see Figure S2.2a in the supplemental material). On the other hand,

Gammaproteobacteria dominated the community of the IF sample in Plant B and was significantly (q<0.05) increased, along with Flavobacteria and Alpha-proteobacteria in the EF sample despite the treatment process. Nonetheless, significant decreases (q<0.05) were seen in majority of the remaining phyla (see Figure S2.2b in the supplemental material).

2.4.2 Bacterial pathogen detection and analysis

Annotation with the MG-RAST pipeline and assignment with the LCA algorithm revealed that at the genus taxonomic level, a total of 20 genera containing potentially pathogenic species were detected across all samples from the list of known bacterial pathogens. Among the genera identified, members belonging to the *Acinetobacter* genus had the highest abundance across all samples, accounting for 11.11% to 81.90% of all pathogenic sequences. This was followed by members of the *Pseudomonas* (6.60% to 15.49%), *Escherichia* (2.76% to 29.85%), *Vibrio* (2.76% to 8.69%), *Clostridium* (1.38% to 14.44%) and *Salmonella* (0.92% to 7.22%) genera (Figure 2.3a).

At the species taxonomic level, sequences assigned to potentially pathogenic species revealed that the bacterial pathogen accounted for a small population in the communities accounting for 0.53% (IF) and 0.46% (EF), and 0.76% (IF) and 0.95% (EF) of the community sequences for Plant A and Plant B, respectively. Here, 355 reads (25 species), 180 reads (29 species), 794 reads (31 species) and 652 reads (23 species) were closely related to known pathogens in Plant A IF, EF and Plant B IF and EF samples, respectively. Amongst the species identified, *Acinetobacter baumannii* was the major pathogenic bacteria in all samples, followed by *Pseudomonas aeruginosa*, *Escherichia coli*, *Vibrio cholera* and *Salmonella enterica* across all

samples (see Table S2.3 in the supplemental material). However, statistical analysis indicated that only *Burkholderia pseudomallei*, *Staphylococcus aureus*, *Helicobacter pylori*, *P. aeruginosa*, *Neisseria meningitides*, *Haemophilus influenzae*, *E. coli*, *A. baumannii*, *S. enterica* and *L. pneumophila* significantly persisted throughout the treatment process in both WWTPs (Figure 2.4a).

As human bacterial pathogens have been well-studied in both the medical and biological fields in terms of its virulence factors (VFs), with these VFs shown to directly cause diseases, they are well suited to serve as indicators for the presence of these bacterial pathogens. Annotation and assignment with the VFDB revealed at the genus taxonomic level, a total of 17 out of the 24 genera of human bacterial pathogens were present across all samples. Among the genera identified, members belonging to the *Pseudomonas* genus had the highest abundance across all samples, accounting for 1.80% to 47.69% of all pathogenic sequences. This was followed by members of the *Legionella* (5.41% to 26.15%), *Neisseria* (7.21% to 15.56%) and *Listeria* (1.81% to 11.11%) genera (Figure 2.3b).

At the species taxonomic level, sequences assigned to potentially pathogenic species revealed that bacterial pathogens accounted for an even smaller population in the communities accounting for 0.067%, 0.069%, 0.11% and 0.096% of the community sequences for Plant A IF, EF and Plant B IF and EF samples, respectively. Here, 45 reads (13 species), 27 reads (11 species), 111 reads (19 species) and 65 reads (8 species) were closely related to known pathogens in Plant A IF, EF and Plant B IF and EF samples, respectively. Amongst the species identified, *P. aeruginosa* was the major pathogenic bacteria in all samples, followed by *L.*

pneumophila, *N. meningitidis*, *L. monocytogenes*, *Brucella melitensis*, *Burkholderia pseudomallei* and *Yersinia enterocolitica* across all samples (see Table S2.4 in the supplemental material). However, statistical analysis indicated that only *B. pseudomallei*, *N. meningitides*, *L. pneumophila*, *P. aeruginosa*, *E. coli* and *L. monocytogenes* significantly persisted throughout the treatment process in both WWTPs (Figure 2.4b).

2.4.3 Composition, abundance and diversity of antibiotic resistance genes

From the communities in the IF received by Plant A, a total of 99 reads were identified as ARGs-like. However, upon treatment EF communities revealed a total of 31 reads determined as ARGs-like. In comparison, in the IF received by Plant B showed a total of 175 reads annotated as ARGs in the communities. However, upon treatment, EF communities showed an increase to 187 reads annotated as ARGs-like. A significant (p<0.05) difference in the total abundance of the detected ARGs was observed between the IF and EF sample in Plant A, but not in Plant B (p>0.05).

Among the 27 ARGs types in the structured CARD database, a total of 14 were detected in both IF and EF samples of Plant A, with 11 types shared between the metagenomic samples. Additionally an antimicrobial peptide resistance gene, *bacA* which recycles undecaprenyl pyrophosphate during cell wall biosynthesis that confers resistance to bacitracin, was also detected in the IF sample for Plant A. For Plant B, a total of 15 types were detected in IF and EF samples, with 12 types shared between the metagenomic samples. The distribution and relative abundance of the different types of ARGs conferring resistance to most major classes of antibiotics were seen across all samples in this study (Figure 2.5). Diversity indices

calculations consistently indicated a reduced diversity of ARGs types in EF communities compared to the initial diversity in IF samples of both WWTPs. Furthermore, it should be noted that the ARGs diversity in the IF sample of Plant B was the highest amongst all wastewater samples. The Shannon diversity of the ARGs types in the EF samples only showed a 19.13% and 4.64% reduction compared to the IF sample diversity for Plant A and B, respectively (see Figure S2.3 in the supplemental material). Nonetheless, PCoA of the ARGs type profiles showed separate clustering between the various wastewater samples (see Figure S2.4 in the supplemental material).

The elfamycin class of ARGs was the dominant type across all metagenomes, accounting for 16.17% to 41.94% of the relative abundance of ARGs, followed by resistance genes for tetracycline (10.86% to 20.20%), aminocoumarin (3.23% to 17.65%), aminoglycoside (6.42% to 14.14%), beta-lactam (6.45% to 11.43%), rifampin (5.05% to 9.68%) and multidrug resistance genes (3.23% to 7.35%) and (Figure 2.5). However, distribution of the resistance genes to various antibiotic classes were not even as indicated by the Buzas and Gibson's evenness index, where the IF communities of both WWTPs showed a higher index compared to the EF communities (see Figure S2.3 in the supplemental material).

The ARGs detected in these water samples were associated with the three major antibiotic resistance mechanisms (extrusion by efflux pumps, antibiotic inactivation or cellular protection). Though, majority belonged to the cellular protection mechanism.

Under the elfamycin ARG class, variants of the elongation factor Tu were the dominant type of ARGs across all samples. Point mutations that occurs in *Mycobacterium tuberculosis* and *Escherichia coli* beta-subunit of RNA polymerase (*rpoB*) conferring resistance to rifampicin was found to be in high abundance in Plant B IF and EF communities. Additionally, genes such as *APH(3'')-Ib*, *tet39*, *adeJ*, *msrE* and *mexT* were also detected in these samples. Similarly, *E. coli rpoB* was also detected in high abundance in Plant A IF, but was successfully reduced in the EF communities. Genes such as *tetQ*, *tetW*, *APH(3'')-Ib*, *msrE*, *sul1*, *tet39* were found to be high abundance in IF and EF samples of Plant A.

2.5 Discussion

In this study, whole metagenome shotgun pyrosequencing combined with metagenomic analysis was used to explore the bacterial community composition, pathogenic bacterial population and associated ARGs in two full-scale urban WWTPs receiving different wastewaters. Direct shotgun pyrosequencing was carried out to avoid the inert bias from 16S rRNA gene targeting PCR, which has previously been shown to miss minor populations in a given sample (Lu et al., 2015). Hence, shotgun sequencing of the metagenomic libraries has allowed for a deeper insight into the complex communities present in the WWTPs influent and effluent samples, with thousands of reads generated, annotated and assigned to different taxa, identified as potential pathogens and varying ARG categories.

2.5.1 Phylogenetic signature of metagenomic sequences

Although there were notable differences in the composition of the bacterial communities between the influent and effluent samples of both WWTPs, Proteobacteria was observed to be the dominant phyla across all samples. This finding is in agreement with several previously reported studies investigating WWTPs using different methods, such as 16S rRNA gene-PCR analysis of community samples collected from WWTPs (Miura et al., 2007; Silva et al., 2010), DNA cloning of activated sludge samples (Snaidr et al., 1997), microarray analysis of biological wastewater treatment reactors (Xia et al., 2010) and several metagenomic analysis of different WWTPs (Ferrera and Sánchez, 2016; Hu et al., 2012; Lee et al., 2015; Sanapareddy et al., 2009; Zhang et al., 2012). As Proteobacteria is known to comprise of one of the most phylogeneticaly and metabolically versatile group in the Bacterial domain (Ettema and Andersson, 2009), considering the setting in which wastewater provides for microbial proliferation, their predominance in such an environment is not surprising. Furthermore, a study which examined the global patterns of bacterial communities from different habitats suggested that Proteobacteria typically occupies an average of 40% of a bacterial population (Nemergut et al., 2011). Hence, as WWTPs present with an ever changing environment, in terms of its nutritional and/or pollutant composition and concentration, the dominance by this phylum across all samples is not surprising. Besides dominance by Proteobacteria, high abundance of several other phyla were also observed across all samples in this study, including members belonging to the Bacteroidetes, Actinobacteria and Firmicutes phyla. Detection of these phyla has been well documented in recent studies investigating WWTPs, such as untreated wastewaters (Shanks et al., 2013), sewage wastewaters (McLellan et al., 2010), swine wastewaters (Da Silva et al., 2015), anaerobic reactor digesting activated sludge from WWTP (Guo et al., 2015), tannery wastewater (Wang et al., 2013) and activated sludge of a WWTP in Hong Kong (Yu and Zhang, 2012). This is in accordance with findings reported by Shanks and

co-workers (Shanks et al., 2013) who investigated the microbial composition of untreated wastewaters at different geographical locations. Furthermore, it has been suggested that the core human microbial signature is composed of members belonging to the Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria phyla (Buffie and Pamer, 2013). Hence, their presence and abundance in the wastewater samples is not surprising as both WWTPs in this study receives influent from domestic sources, with Plant B receiving a further mix of nearby hospital effluents. Hence, this suggests that at this taxonomic level of complexity, though significant changes are observed upon treatment, it does not result in the overall change of the bacterial community signature of wastewaters being discharged into the environment.

Further phylogenetic analysis at the class taxonomic level revealed that members belonging to the Beta-, Epsilon-, Gamma-proteobacteria, Bacteroidia, Clostridia, Flavobacteria and Actinobacteria dominated the IF wastewaters received by both plants. Ye and Zhang (2013) reported Delta-, Gamma-proteobacteria, and Clostridia as the dominant classes in the influent wastewaters analysed, whilst McLellan *et al.* (2010) reported WWTP influent dominated by Actinobacteria, Bacteroidetes, and Firmicutes classes. These differences may be attributed to many factors, such as the composition of the wastewater received by both plants, the spatial and temporal variations, nearby industry effluents discharged and the treatment technology adopted by the WWTPs (McLellan et al., 2010). Additionally, fluctuations between members of these classes observed when IF and EF samples were compared to each other, further suggesting the contribution of wastewater treatment in the changes to the bacterial community structure.

2.5.2 Pathogenic bacterial populations of untreated and treated wastewaters

In the MG-RAST annotations, a relatively standard cut-off was employed to filter the BLAST outputs. Furthermore, we employed very strict cutoffs for the VFDB annotations. For this reason, discrepancies between the detected potential pathogens was observed between the output of both analysis, whereby a reduced number of pathogenic genera was detected in the VFDB outputs. Hence, comparison of the potential pathogenic genera showed that both databases only shared *Campylobacter*, *Escherichia*, *Haemophilus*, *Helicobacter*, *Legionella*, *Listeria*, *Mycobacterium*, *Neisseria*, *Pseudomonas*, *Salmonella*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Vibrio* and *Yersinia* genera. Nonetheless, annotation by both M5nr database and VFDB consistently showed that the abundance and diversity of the observed potential bacterial pathogens declined as a result of the treatment process, from the IF along the treatment train to the EF. Previous studies have demonstrated a removal efficiency of over 99% for 12 pathogens, including *A. hydrophila*, *K. pneumoniae* and *E. coli*, with the use of the activated sludge process detected by microarrays and quantitative-PCR (Lee et al., 2006). Therefore, since WWTPs in this study applied this technology during the treatment process, it suggests that the treatment process may be effective in the reduction of the pathogenic load by both WWTPs before the subsequent release of the wastewaters into the aquatic environments. However, since 8 to 19 genera containing potential pathogens were still detected from in EF samples of both WWTPs, with some observed at a relatively high occurrence, there are concerns about the safety of the treated effluent discharge from these WWTPs on the receiving environment.

Acinetobacter baumannii strains have previously been reported to be present in wastewaters receiving hospital discharges (Ferreira et al., 2011; Zhang et al., 2014) and only a few studies have reported their detection elsewhere in the environment (El-Sayed, 2016; Girlich et al., 2010;

Hrenovic et al., 2014). Surprisingly, a 4-fold increase of *A. baumannii* in treated wastewater communities of Plant B was observed in this study from the MG-RAST pipeline but not the VFDB analysis. The absence could be explained by the fact that *A. baumannii* virulence factors were not included within the VFDB. The bacterium has been reported as one of the most important human pathogen to cause nosocomial pneumonia and bacteremia among patients admitted at the intensive care unit (Gaynes and Edwards, 2005; Kanafani et al., 2003; Paul et al., 2005; Wisplinghoff et al., 2004) followed by skin, soft tissue, and urinary tract infections (Bergogne-Bérézin and Towner, 1996; Gales et al., 2001) and secondary meningitis (Falagas et al., 2007; Ng et al., 2006) globally. Furthermore, the number of reports of multidrug resistant *A. baumannii* in hospitals has been steadily increasing (Dijkshoorn et al., 2007). Though *A. baumannii* is an ubiquitous organism, as it could be found in various environmental sources such as soil, water, vegetables, meat, and fish (Krahn et al., 2016), they may infrequently colonize the skin of healthy human tissue, although the occurrence is typically at a low-density and for short-term duration (Cetin et al., 2009). Furthermore, as studies have demonstrated the presence of antibiotic resistant *A. baumannii* in a municipal WWTP (Hrenovic et al., 2016) and the contribution of the wastewater treatment process in increased selection of multidrug resistant *Acinetobacter* spp. discharged into the environment (Zhang et al., 2009), the relatively high occurrence of this species in the EF communities warrants further investigation.

Legionella pneumophila is the causative agent of Legionnaires' disease, a type of atypical pneumonia with a relatively high fatality rate, or Pontiac fever, a milder non-fatal form of Legionella infection. Normally, *Legionella* is transmitted mainly by inhaling bio-aerosols of contaminated water (Devos et al., 2005). Overall, approximately a 2.5-fold to 5-fold increase of *L. pneumophila* was observed in the treated wastewater communities of both WWTPs in this

study from both the MG-RAST pipeline and VFDB analysis. Though the organism is known to be ubiquitous in aqueous environments, such as: lakes, rivers, reservoirs, cooling towers, and whirlpools (Dusserre et al., 2008), it has been shown to survive in extreme ranges of environmental conditions, with a variety of different physiochemical factors (Mirzaee et al., 2015). However, it typically requires free-living amoebae for its intracellular replication (Albert-Weissenberger et al., 2007), though, under appropriate conditions, *L. pneumophila* has been reported to survive for long periods as a free organism in low-nutrient environments (Chang et al., 2007; Steinert et al., 1997). Not many cases of *L. pneumophila* diseases have been frequently discussed in literature compared to other bacterial pathogens, but infections caused by *L. pneumophila* has been largely associated with biological treatment plants. For example, a survey of 33 industrial plants in Norway indicated aeration basins as sources for high concentrations of *Legionella* spp. (Lund et al., 2014), whilst two cases of severe pneumonia in employees working at two separate industrial WWTPs was documented in Finland (Kusnetsov et al., 2010). In the latter study, up to 1.7×10^{10} cells/L of *L. pneumophila* was detected in the aeration ponds of the WWTP and 10^5 CFU/L was detected downstream of wastewater outlet. Such levels of *L. pneumophila* was shown to directly cause the outbreak of Legionnaires' disease. Furthermore, some commonly detected genera such as *Escherichia*, *Listeria*, *Neisseria*, *Pseudomonas*, *Salmonella*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Vibrio* and *Yersinia* were also found in the EF of both WWTPs investigated in this study.

2.5.3 Profiles of putative antibiotic resistance gene sequences across metagenomes

The strict and unambiguous antibiotic resistance ontology of the CARD database was used to annotate genes as ARG-like in this study. It should be noted that the diversity and abundance of the antibiotic resistome observed by CARD alignment were different from those derived

from the MG-RAST analysis of the metagenome. Possible reasons for the divergence may be the difference in the reference database used as well as the contribution of the alignment algorithm. MG-RAST relies on the use of BLAT algorithm for similarity searches, which is less sensitive than the BLAST algorithm (Yu and Zhang, 2013). Additionally, the CARD database is a highly curated and comprehensive database (McArthur et al., 2013) compared to the incomplete resistance to antibiotics and toxic compounds subsystem on ARGs (Wang et al., 2013).

Overall, detection frequencies and diversity of different ARGs and ARGs types in the metagenomic samples of both WWTPs were expected as described in previous studies (Shi et al., 2013; Szczepanowski et al., 2009; Wang et al., 2013; Xi et al., 2009). Plant A showed a significant removal of ARGs types following treatment with a lower diversity of ARGs types in the EF communities. This is in line with a recent study by Chen and Zhang (2013) who investigated tetracycline and sulfonamide resistance genes in 4 municipal wastewater and 8 rural domestic sewage treatment systems. Conversely, Plant B communities was observed without salient removal of ARGs types upon wastewater treatment. However, slight differences were observed by all means between the samples. Notably, this result corroborates with the suggestion that WWTPs represent the main sites through which ARGs are released into the environment (Marti et al., 2014).

Genes conferring resistance to the elfamycin, tetracycline and aminoglycoside types of antibiotics was observed to be the most abundant across all metagenomes. The elfamycin family of antibiotics inhibits protein synthesis of Gram-negative bacteria by an interaction with

elongation factor Tu (Hall et al., 1989; Vogeley et al., 2001; Wolf et al., 1974). The tetracycline family of antibiotics also inhibits protein synthesis by preventing the attachment of aminoacyl-tRNA to the ribosomal acceptor A site (Chopra and Roberts, 2001). The aminoglycoside family of antibiotics also inhibits protein synthesis of most Gram-negative aerobic and facultative anaerobic bacilli by the irreversible binding to the 30S ribosomal sub-unit (Kotra et al., 2000). Specifically, majority of the elfamycin-like ARGs in the metagenomes were associated with the variants of the elongation factor Tu. Elfamycins and aminoglycosides are not commonly used in treatment of human diseases due to the very narrow spectrum against human pathogens (Miele et al., 1994) and the adverse effects leading to ototoxicity, neuropathy and nephrotoxicity (Prayle et al., 2010). However, they are applied as a growth promoting agent in animal food production owing to their excellent activities (Brötz et al., 2011; Eagar et al., 2012; Maiese et al., 1989). Therefore, high abundance of these ARGs in the metagenomes suggests that the major source of wastewaters received by both WWTPs could be from nearby animal feed industries around both WWTPs. This reason could also partially explain the coinciding high occurrence of genes conferring resistance to tetracycline class of antibiotics as tetracycline is commonly used for livestock purposes in South Africa (Eagar et al., 2012). On the other hand, studies suggests that sources of *tet39* gene could result from clinical (Agersø and Guardabassi, 2005) or animal sources (Agersø and Petersen, 2007) and are common amongst *Acinetobacter* spp. (Hamidian et al., 2016), which was detected in high abundance in this study. The occurrence of these particular family of antibiotic and their ineffective biodegradation in WWTPs may allow for low concentrations to persist in the treated effluents leading to the selection of these ARGs (Gullberg et al., 2011) and their subsequent dissemination into the environment.

Nonetheless, the proportions of the ARGs identified in this study were diverse and comparable to the results from previous studies investigating activated sludge (Zhang et al., 2011), sewage effluent (Port et al., 2012), plasmids recovered from WWTP (Li et al., 2015a), non-hospital medical care facility (Bäumlisberger et al., 2015) and drinking water treatment plants (Chao et al., 2013; Huang et al., 2011; Xi et al., 2009). Moreover, the ARGs detected in this study encompasses the three major resistance mechanisms, *viz.* exclusion by efflux pumps, antibiotic inactivation and cellular protection (Blair et al., 2014; Nikaido, 2009). Hence, detection of these ARGs in the EF of both WWTPs further demonstrates that WWTPs are hotspots for the ARGs transfer between bacteria and potentially contribute to the increase in the multidrug resistance strains.

2.6 Conclusions

In conclusion, while the sequencing depth in this study was not adequate due to the sequencing platform used, the fact that direct shotgun pyrosequencing combined with metagenomic analysis allowed for the determination of the microbial community structure, potential pathogenic bacterial populations as well as the ARGs profiles of two urban WWTPs in this study supports this approach as a relatively simple yet advantageous alternative over the time-consuming and laborious conventional techniques used in community assessments, pathogen and ARGs detection. The findings of this study exhibited the taxonomic profiles of the wastewater communities, which were in general agreement with previous descriptions of aquatic and wastewater sources reported in literature. Furthermore, the observed profile of commonly concerning human bacterial pathogens was also comprehensively explored in a

single detection step, with *Acinetobacter baumannii* and *Legionella pneumophila* observed in high abundance regardless of treatment and is of concern. Resistance determinants also correlated with previous studies investigating similar environmental sources and our study further indicates the contribution of urban WWTPs in antibiotic resistance generation. Additionally, our study suggests, for the first time, the dominance of the elfamycin family of antibiotics in two different urban WWTPs although the wastewaters were subjected to the complete treatment process. These findings demonstrated the applicability of this technique as a powerful yet feasible approach which could be applied in microbial ecology and other related fields. However, it is worth noting that such a molecular technique is difficult to absolutely quantify the pathogens in question, in terms of an exact cell number, because of the complexity in converting gene copy number to cell number.

2.7 References

Agersø, Y., Guardabassi, L., 2005. Identification of Tet 39, a novel class of tetracycline resistance determinant in *Acinetobacter* spp. of environmental and clinical origin. J. Antimicrob. Chemother. 55, 566–569. doi:10.1093/jac/dki051

Agersø, Y., Petersen, A., 2007. The tetracycline resistance determinant Tet 39 and the sulphonamide resistance gene sulII are common among resistant *Acinetobacter* spp. isolated from integrated fish farms in Thailand. J. Antimicrob. Chemother. 59, 23–27. doi:10.1093/jac/dkl419

Albert-Weissenberger, C., Cazalet, C., Buchrieser, C., 2007. *Legionella pneumophila* - a human pathogen that co-evolved with fresh water protozoa. Cell. Mol. Life Sci. 64, 432–448. doi:10.1007/s00018-006-6391-1

Amos, G.C.A., Zhang, L., Hawkey, P.M., Gaze, W.H., Wellington, E.M., 2014. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet. Microbiol. 171, 441–447. doi:10.1016/j.vetmic.2014.02.017

Aw, T.G., Rose, J.B., 2012. Detection of pathogens in water: from phylochips to qPCR to pyrosequencing. Curr. Opin. Biotechnol. 23, 422–430. doi:10.1016/j.copbio.2011.11.016

Bäumlisberger, M., Youssar, L., Schilhabel, M.B., Jonas, D., 2015. Influence of a non-hospital medical care facility on antimicrobial resistance in wastewater. PLoS One 10, e0122635. doi:10.1371/journal.pone.0122635

Bergogne-Bérézin, E., Towner, K.J., 1996. *Acinetobacter* spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features. Clin. Microbiol. Rev. 9, 148–165.

Blair, J.M.A., Webber, M.A., Baylay, A.J., Ogbolu, D.O., Piddock, L.J. V., 2014. Molecular mechanisms of antibiotic resistance. Nat. Rev. Microbiol. 13, 42–51. doi:10.1038/nrmicro3380

Brötz, E., Kulik, A., Vikineswary, S., Lim, C.-T., Tan, G.Y.A., Zinecker, H., Imhoff, J.F., Paululat, T., Fiedler, H.-P., 2011. Phenelfamycins G and H, new elfamycin-type antibiotics produced by *Streptomyces albospinus* Acta 3619. J. Antibiot. (Tokyo). 64, 257–66. doi:10.1038/ja.2010.170

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. doi:10.1038/nmeth.3176

Buffie, C.G., Pamer, E.G., 2013. Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol. 13, 790–801. doi:10.1038/nri3535

Cao, B., Li, R., Xiong, S., Yao, F., Liu, X., Wang, M., Feng, L., Wang, L., 2011. Use of a DNA microarray for detection and identification of bacterial pathogens associated with fishery products. Appl. Environ. Microbiol. 77, 8219–8225. doi:10.1128/AEM.05914-11

Cetin, E.S., Durmaz, R., Tetik, T., Otlu, B., Kaya, S., Calişkan, A., 2009. Epidemiologic characterization of nosocomial *Acinetobacter baumannii* infections in a Turkish university hospital by pulsed-field gel electrophoresis. Am. J. Infect. Control 37, 56–64. doi:10.1016/j.ajic.2008.01.010

Chang, C.-W., Hwang, Y.-H., Cheng, W.-Y., Chang, C.-P., 2007. Effects of chlorination and heat disinfection on long-term starved *Legionella pneumophila* in warm water. J. Appl. Microbiol. 102, 1636–1644. doi:10.1111/j.1365-2672.2006.03195.x

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X.-X., Wu, W.-M., Zhang, T., 2013. Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci. Rep. 3, 3550. doi:10.1038/srep03550

Chen, H., Zhang, M., 2013. Occurrence and removal of antibiotic resistance genes in municipal wastewater and rural domestic sewage treatment systems in eastern China. Environ. Int. 55, 9–14. doi:10.1016/j.envint.2013.01.019

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., Jin, Q., 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 33, D325–D328. doi:10.1093/nar/gki008

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2015. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 44, D694–D697. doi:10.1093/nar/gkv1239

Chopra, I., Roberts, M., 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol. Mol. Biol. Rev. 65, 232–260. doi:10.1128/MMBR.65.2.232-260.2001

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485. doi:10.1186/1471-2105-11-485

Da Silva, M.L.B., Cantão, M.E., Mezzari, M.P., Ma, J., Nossa, C.W., 2015. Assessment of bacterial and archaeal community structure in Swine wastewater treatment processes. Microb. Ecol. 70, 77–87. doi:10.1007/s00248-014-0537-8

Department of Water and Sanitation, 2015. Green Drop Progress Report. URL https://www.dwa.gov.za/dir_ws/gds/Docs/DocsDefault.aspx (accessed 4.28.16).

Devos, L., Boon, N., Verstraete, W., 2005. *Legionella pneumophila* in the Environment: the occurrence of a fastidious bacterium in oligotrophic conditions. Rev. Environ. Sci. Bio/Technology 4, 61–74. doi:10.1007/s11157-004-8174-1

Dijkshoorn, L., Nemec, A., Seifert, H., 2007. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. Nat. Rev. Microbiol. 5, 939–951. doi:10.1038/nrmicro1789

Doxey, A.C., Kurtz, D.A., Lynch, M.D.J., Sauder, L.A., Neufeld, J.D., 2014. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. ISME J. 9, 461–471. doi:10.1038/ismej.2014.142

Dusserre, E., Ginevra, C., Hallier-Soulier, S., Vandenesch, F., Festoc, G., Etienne, J., Jarraud, S., Molmeret, M., 2008. A PCR-based method for monitoring *Legionella pneumophila* in water samples detects viable but noncultivable legionellae that can recover their cultivability. Appl. Environ. Microbiol. 74, 4817–4824. doi:10.1128/AEM.02899-07

Eagar, H., Swan, G., van Vuuren, M., 2012. A survey of antimicrobial usage in animals in South Africa with specific reference to food animals. J. S. Afr. Vet. Assoc. 83, 16. doi:10.4102/jsava.v83i1.16

El-Sayed, M.H., 2016. Multiple Heavy Metal and Antibiotic Resistance of *Acinetobacter baumannii* Strain HAF – 13 Isolated from Industrial Effluents. Am. J. Microbiol. Res. 4, 26–36. doi:10.12691/ajmr-4-1-3

Ettema, T.J.G., Andersson, S.G.E., 2009. The alpha-proteobacteria: the Darwin finches of the bacterial world. Biol. Lett. 5, 429–432. doi:10.1098/rsbl.2008.0793

Falagas, M.E., Bliziotis, I.A., Tam, V.H., 2007. Intraventricular or intrathecal use of polymyxins in patients with Gram-negative meningitis: a systematic review of the available evidence. Int. J. Antimicrob. Agents 29, 9–25. doi:10.1016/j.ijantimicag.2006.08.024

Ferreira, A.E., Marchetti, D.P., De Oliveira, L.M., Gusatti, C.S., Fuentefria, D.B., Corção, G., 2011. Presence of OXA-23-producing isolates of *Acinetobacter baumannii* in wastewater from hospitals in southern Brazil. Microb. Drug Resist. 17, 221–227. doi:10.1089/mdr.2010.0013

Ferrera, I., Sánchez, O., 2016. Insights into microbial diversity in wastewater treatment systems: How far have we come? Biotechnol. Adv. in press. doi:10.1016/j.biotechadv.2016.04.003

Gales, A.C., Jones, R.N., Forward, K.R., Liñares, J., Sader, H.S., Verhoef, J., 2001. Emerging importance of multidrug-resistant *Acinetobacter* species and *Stenotrophomonas maltophilia* as pathogens in seriously ill patients: geographic patterns, epidemiological features, and trends in the SENTRY Antimicrobial Surveillance Program (1997-1999. Clin. Infect. Dis. 32, S104–S113. doi:10.1086/320183

Gaynes, R., Edwards, J.R., 2005. Overview of nosocomial infections caused by Gram-negative bacilli. Clin. Infect. Dis. 41, 848–854. doi:10.1086/432803

Gaze, W.H., Zhang, L., Abdouslam, N.A., Hawkey, P.M., Calvo-Bado, L., Royle, J., Brown, H., Davis, S., Kay, P., Boxall, A.B.A., Wellington, E.M.H., 2011. Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. ISME J. 5, 1253–1261. doi:10.1038/ismej.2011.15

Girlich, D., Poirel, L., Nordmann, P., 2010. First isolation of the *blaOXA-23* carbapenemase gene from an environmental *Acinetobacter baumannii* isolate. Antimicrob. Agents Chemother. 54, 578–579. doi:10.1128/AAC.00861-09

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317. doi:10.1038/ismej.2009.72

Gullberg, E., Cao, S., Berg, O.G., Ilbäck, C., Sandegren, L., Hughes, D., Andersson, D.I., 2011. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog. 7, e1002158. doi:10.1371/journal.ppat.1002158

Guo, J., Peng, Y., Ni, B.-J., Han, X., Fan, L., Yuan, Z., 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. Microb. Cell Fact. 14, 33. doi:10.1186/s12934-015-0218-4

Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L., Rolain, J.-M., 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob. Agents Chemother. 58, 212–220. doi:10.1128/AAC.01310-13

Hall, C.C., Watkins, J.D., Georgopapadakou, N.H., 1989. Effects of elfamycins on elongation factor Tu from *Escherichia coli* and *Staphylococcus aureus*. Antimicrob. Agents Chemother. 33, 322–325.

Hamidian, M., Holt, K.E., Pickard, D., Hall, R.M., 2016. A small *Acinetobacter* plasmid carrying the *tet39* tetracycline resistance determinant. J. Antimicrob. Chemother. 71, 269–271. doi:10.1093/jac/dkv293

Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4, 1–9.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M., Brown, C.T., 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proc. Natl. Acad. Sci. 111, 4904–4909. doi:10.1073/pnas.1402564111

Hrenovic, J., Durn, G., Goic-Barisic, I., Kovacic, A., 2014. Occurrence of an environmental *Acinetobacter baumannii* strain similar to a clinical isolate in paleosol from Croatia. Appl. Environ. Microbiol. 80, 2860–2866. doi:10.1128/AEM.00312-14

Hrenovic, J., Goic-Barisic, I., Kazazic, S., Kovacic, A., Ganjto, M., Tonkic, M., 2016. Carbapenem-resistant isolates of *Acinetobacter baumannii* in a municipal wastewater treatment plant, Croatia, 2014. Euro Surveill. 21, 30195. doi:http://dx.doi.org/10.2807/1560-7917.ES.2016.21.15.30195

Hu, M., Wang, X., Wen, X., Xia, Y., 2012. Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. Bioresour. Technol. 117, 72–79. doi:10.1016/j.biortech.2012.04.061

Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., Pan, Y., Li, J., Zhu, L., Wang, X., Meng, Z., Zhao, F., Liu, D., Ma, J., Qin, N., Xiang, C., Xiao, Y., Li, L., Yang, H., Wang, J., Yang, R., Gao, G.F., Wang, J., Zhu, B., 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat. Commun. 4, 2151. doi:10.1038/ncomms3151

Huang, J.-J., Hu, H.-Y., Tang, F., Li, Y., Lu, S.-Q., Lu, Y., 2011. Inactivation and reactivation of antibiotic-resistant bacteria by chlorination in secondary effluents of a municipal wastewater treatment plant. Water Res. 45, 2775–2781. doi:10.1016/j.watres.2011.02.026

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. doi:10.1101/gr.5969107

Ilmberger, N., Güllert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., Alawi, M., Poehlein, A., Chow, J., Turaev, D., Rattei, T., Schmeisser, C., Salomon, J., Olsen, P.B., Daniel, R., Grundhoff, A., Borchert, M.S., Streit, W.R., 2014. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One 9, e106707. doi:10.1371/journal.pone.0106707

Johnning, A., Moore, E.R.B., Svensson-Stadler, L., Shouche, Y.S., Larsson, D.G.J., Kristiansson, E., 2013. Acquired Genetic Mechanisms of a Multiresistant Bacterium Isolated from a Treatment Plant Receiving Wastewater from Antibiotic Production. Appl. Environ. Microbiol. 79, 7256–7263. doi:10.1128/AEM.02141-13

Ju, F., Guo, F., Ye, L., Xia, Y., Zhang, T., 2014. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. Environ. Microbiol. Rep. 6, 80–89. doi:10.1111/1758-2229.12110

Kanafani, Z.A., Kara, L., Hayek, S., Kanj, S.S., 2003. Ventilator-associated pneumonia at a tertiary-care center in a developing country: incidence, microbiology, and susceptibility patterns of isolated microorganisms. Infect. Control Hosp. Epidemiol. 24, 864–869. doi:10.1086/502151

Kim, J., Lim, J., Lee, C., 2013. Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations. Biotechnol. Adv. 31, 1358–1373. doi:10.1016/j.biotechadv.2013.05.010

Kotra, L.P., Haddad, J., Mobashery, S., 2000. Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance. Antimicrob. Agents Chemother. 44, 3249–3256. doi:10.1128/AAC.44.12.3249-3256.2000

Krahn, T., Wibberg, D., Maus, I., Winkler, A., Bontron, S., Sczyrba, A., Nordmann, P., Pühler, A., Poirel, L., Schlüter, A., 2016. Intraspecies transfer of the chromosomal *Acinetobacter baumannii blaNDM-1* carbapenemase gene. Antimicrob. Agents Chemother. 60, 3032–3040. doi:10.1128/AAC.00124-16

Kristiansson, E., Fick, J., Janzon, A., Grabic, R., Rutgersson, C., Weijdegård, B., Söderström, H., Larsson, D.G.J., 2011. Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One 6, e17038. doi:10.1371/journal.pone.0017038

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19, 1639–1645. doi:10.1101/gr.092759.109

Kusnetsov, J., Neuvonen, L.-K., Korpio, T., Uldum, S.A., Mentula, S., Putus, T., Tran Minh, N.N., Martimo, K.-P., 2010. Two Legionnaires' disease cases associated with industrial waste water treatment plants: a case report. BMC Infect. Dis. 10, 343. doi:10.1186/1471-2334-10-343

Lee, D.-Y., Shannon, K., Beaudette, L.A., 2006. Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. J. Microbiol. Methods 65, 453–467. doi:10.1016/j.mimet.2005.09.008

Lee, S.-H., Kang, H.-J., Park, H.-D., 2015. Influence of influent wastewater communities on temporal variation of activated sludge communities. Water Res. 73, 132–244. doi:10.1016/j.watres.2015.01.014

Li, A.-D., Li, L.-G., Zhang, T., 2015a. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. Front. Microbiol. 6, 1025. doi:10.3389/fmicb.2015.01025

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., Zhang, T., 2015b. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J. 9, 2490–2502. doi:10.1038/ismej.2015.59

Li, D., Yu, T., Zhang, Y., Yang, M., Li, Z., Liu, M., Qi, R., 2010. Antibiotic resistance characteristics of environmental bacteria from an oxytetracycline production wastewater treatment plant and the receiving river. Appl. Environ. Microbiol. 76, 3444–3451. doi:10.1128/AEM.02964-09

Lim, Y.W., Evangelista, J.S., Schmieder, R., Bailey, B., Haynes, M., Furlan, M., Maughan, H., Edwards, R., Rohwer, F., Conrad, D., 2014. clinical insights from metagenomic analysis of sputum samples from patients with Cystic Fibrosis. J. Clin. Microbiol. 52, 425–437. doi:10.1128/JCM.02204-13

Liu, B., Pop, M., 2009. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res. 37, D443–D447. doi:10.1093/nar/gkn656

Luo, Y., Mao, D., Rysz, M., Zhou, Q., Zhang, H., Xu, L., J. J. Alvarez, P., 2010. Trends in Antibiotic Resistance Genes Occurrence in the Haihe River, China. Environ. Sci. Technol. 44, 7220–7225. doi:10.1021/es100233w

Lu, X., Zhang, X.-X., Wang, Z., Huang, K., Wang, Y., Liang, W., Tan, Y., Liu, B., Tang, J., 2015. Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. PLoS One 10, e0125549. doi:10.1371/journal.pone.0125549

Lund, V., Fonahn, W., Pettersen, J.E., Caugant, D.A., Ask, E., Nysaeter, A., 2014. Detection of Legionella by cultivation and quantitative real-time polymerase chain reaction in biological waste water treatment plants in Norway. J. Water Health 12, 543–554. doi:10.2166/wh.2014.063

Maiese, W.M., Lechevalier, M.P., Lechevalier, H.A., Korshalla, J., Goodman, J., Wildey, M.J., Kuck, N., Conner, S.D., Greenstein, M., 1989. LL-E19020 alpha and beta, animal growth promoting antibiotics: taxonomy, fermentation and biological activity. J. Antibiot. 42, 1489–1493.

Marti, E., Variatza, E., Balcazar, J.L., 2014. The role of aquatic ecosystems as reservoirs of antibiotic resistance. Trends Microbiol. 22, 36–41. doi:10.1016/j.tim.2013.11.001

Mason, O.U., Scott, N.M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., Bouskill, N.J., Prestat, E., Borglin, S., Joyner, D.C., Fortney, J.L., Jurelevicius, D., Stringfellow, W.T., Alvarez-Cohen, L., Hazen, T.C., Knight, R., Gilbert, J.A., Jansson, J.K., 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J. 8, 1464–1475. doi:10.1038/ismej.2013.254

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., Pascale, G. De, Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J. V, Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., Wright, G.D., 2013. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57, 3348–3357. doi:10.1128/AAC.00419-13

McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N., Sogin, M.L., 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ. Microbiol. 12, 378–392. doi:10.1111/j.1462-2920.2009.02075.x

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386. doi:10.1186/1471-2105-9-386

Miele, A., Goldstein, B.P., Bandera, M., Jarvis, C., Resconi, A., Williams, R.J., 1994. Differential susceptibilities of enterococcal species to elfamycin antibiotics. J. Clin. Microbiol. 32, 2016–2018.

Mirzaee, S.A., Nikaeen, M., Hajizadeh, Y., Nabavi, B.F., Hassanzadeh, A., 2015. Detection of *Legionella* spp. by a nested-PCR assay in air samples of a wastewater treatment plant and downwind distances in Isfahan. Adv. Biomed. Res. 4, 48. doi:10.4103/2277-9175.151540

Miura, Y., Hiraiwa, M.N., Ito, T., Itonaga, T., Watanabe, Y., Okabe, S., 2007. Bacterial community structures in MBRs treating municipal wastewater: relationship between community stability and reactor performance. Water Res. 41, 627–637. doi:10.1016/j.watres.2006.11.005

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L., Knight, R., 2011. Global patterns in the biogeography of bacterial taxa. Environ. Microbiol. 13, 135–144. doi:10.1111/j.1462-2920.2010.02315.x

Ng, J., Gosbell, I.B., Kelly, J.A., Boyle, M.J., Ferguson, J.K., 2006. Cure of multiresistant *Acinetobacter baumannii* central nervous system infections with intraventricular or intrathecal colistin: case series and literature review. J. Antimicrob. Chemother. 58, 1078–1081. doi:10.1093/jac/dkl347

Nikaido, H., 2009. Multidrug resistance in bacteria. Annu. Rev. Biochem. 78, 119–146. doi:10.1146/annurev.biochem.78.082907.145923

Okoh, A., Odjadjare, E., Igbinosa, E., Osode, A., 2007. Wastewater treatment plants as a source of microbial pathogens in receiving watersheds. African J. Biotechnol. doi:10.4314/ajb.v6i25.58260

Pacchioni, R.G., Carvalho, F.M., Thompson, C.E., Faustino, A.L.F., Nicolini, F., Pereira, T.S., Silva, R.C.B., Cantão, M.E., Gerber, A., Vasconcelos, A.T.R., Agnez-Lima, L.F., 2014. Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil. Microbiologyopen 3, 299–315. doi:10.1002/mbo3.169

Parks, D.H., Beiko, R.G., 2010. Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26, 715–721. doi:10.1093/bioinformatics/btq041

Paul, M., Weinberger, M., Siegman-Igra, Y., Lazarovitch, T., Ostfeld, I., Boldur, I., Samra, Z., Shula, H., Carmeli, Y., Rubinovitch, B., Pitlik, S., 2005. *Acinetobacter baumannii*: emergence and spread in Israeli hospitals 1997-2002. J. Hosp. Infect. 60, 256–260. doi:10.1016/j.jhin.2005.01.007

Port, J.A., Wallace, J.C., Griffith, W.C., Faustman, E.M., 2012. Metagenomic profiling of microbial composition and antibiotic resistance determinants in Puget Sound. PLoS One 7, e48000. doi:10.1371/journal.pone.0048000

Prayle, A., Watson, A., Fortnum, H., Smyth, A., 2010. Side effects of aminoglycosides on the kidney, ear and balance in cystic fibrosis. Thorax 65, 654–658. doi:10.1136/thx.2009.131532

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K. Le, Leclerc, M., Maguin,

E., Minardi, R.M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65. doi:10.1038/nature08821

R Development Core Team, 2007. R: A Language and Environment for Statistical Computing.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I., Fatta-Kassinos, D., 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci. Total Environ. 447, 345–360. doi:10.1016/j.scitotenv.2013.01.032

Sanapareddy, N., Hamp, T.J., Gonzalez, L.C., Hilger, H.A., Fodor, A.A., Clinton, S.M., 2009. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Appl. Environ. Microbiol. 75, 1688–1696. doi:10.1128/AEM.01210-08

Scaria, J., Chandramouli, U., Verma, S.K., 2005. Antibiotic Resistance Genes Online (ARGO): a database on vancomycin and beta-lactam resistance genes. Bioinformation 1, 5–7.

Schmieder, R., Edwards, R., 2012. Insights into antibiotic resistance through metagenomic approaches. Future Microbiol. 7, 73–89. doi:10.2217/fmb.11.135

Schwartz, T., Kohnen, W., Jansen, B., Obst, U., 2003. Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water biofilms. FEMS Microbiol. Ecol. 43, 325–335. doi:10.1111/j.1574-6941.2003.tb01073.x

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl. Environ. Microbiol. 79, 2906–2913.

Shen, S.Y., Fulthorpe, R., 2015. Seasonal variation of bacterial endophytes in urban trees. Front. Microbiol. 6, 427. doi:10.3389/fmicb.2015.00427

Shi, P., Jia, S., Zhang, X.-X., Zhang, T., Cheng, S., Li, A., 2013. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res. 47, 111–120. doi:10.1016/j.watres.2012.09.046

Silva, C.C., Jesus, E.C., Torres, A.P.R., Sousa, M.P., Santiago, V.M.J., Oliveira, V.M., 2010. Investigation of bacterial diversity in membrane bioreactor and conventional activated sludge processes from petroleum refineries using phylogenetic and statistical approaches. J. Microbiol. Biotechnol. 20, 447–459.

Snaidr, J., Amann, R., Huber, I., Ludwig, W., Schleifer, K.H., 1997. Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. Appl. Environ. Microbiol. 63, 2884–2896.

Steinert, M., Emödy, L., Amann, R., Hacker, J., 1997. Resuscitation of viable but nonculturable *Legionella pneumophila* Philadelphia JR32 by *Acanthamoeba castellanii*. Appl. Environ. Microbiol. 63, 2047–2053.

Storteboom, H., Arabi, M., Davis, J.G., Crimi, B., Pruden, A., 2010. Identification of antibiotic-resistance-gene molecular signatures suitable as tracers of pristine river, urban, and agricultural sources. Environ. Sci. Technol. 44, 1947–1953. doi:10.1021/es902893f

Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K.-H., Gützkow, T., Eichler, W., Pühler, A., Schlüter, A., 2009. Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. Microbiology 155, 2306–2319. doi:10.1099/mic.0.028233-0

Toze, S., 1999. PCR and the detection of microbial pathogens in water and wastewater. Water Res. 33, 3545–3556. doi:10.1016/S0043-1354(99)00071-8

Vogeley, L., Palm, G.J., Mesters, J.R., Hilgenfeld, R., 2001. Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. J. Biol. Chem. 276, 17149–17155. doi:10.1074/jbc.M100017200

Wang, Z., Zhang, X.-X., Huang, K., Miao, Y., Shi, P., Liu, B., Long, C., Li, A., 2013. Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. PLoS One 8, e76079. doi:10.1371/journal.pone.0076079

Wen, Q., Tutuka, C., Keegan, A., Jin, B., 2009. Fate of pathogenic microorganisms and indicators in secondary activated sludge wastewater treatment plants. J. Environ. Manage. 90, 1442–1447. doi:10.1016/j.jenvman.2008.09.002

White, R.A., Power, I.M., Dipple, G.M., Southam, G., Suttle, C.A., 2015. Metagenomic analysis reveals that modern microbialites and polar microbial mats have similar taxonomic and functional potential. Front. Microbiol. 6, 966. doi:10.3389/fmicb.2015.00966

WHO, 2014. Antimicrobial resistance: global report on surveillance 2014. URL http://www.who.int/drugresistance/documents/surveillancereport/en/ (accessed 4.28.16).

WHO, 2006. Guidelines for the safe use of wastewater, excreta and greywater - Policy and regulatory aspects. URL http://www.who.int/water_sanitation_health/wastewater/gsuww/en/ (accessed 4.28.16).

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P., Paczian, T., Trimble, W.L., Bagchi, S., Grama, A., Chaterji, S., Meyer, F., 2015. The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Res. 44, D590–D594. doi:10.1093/nar/gkv1322

Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P., Edmond, M.B., 2004. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin. Infect. Dis. 39, 309–317. doi:10.1086/421946

Wolf, H., Chinali, G., Parmeggiani, A., 1974. Kirromycin, an inhibitor of protein biosynthesis that acts on elongation factor Tu. Proc. Natl. Acad. Sci. U. S. A. 71, 4910–4914.

Xi, C., Zhang, Y., Marrs, C.F., Ye, W., Simon, C., Foxman, B., Nriagu, J., 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Appl. Environ. Microbiol. 75, 5714–5718. doi:10.1128/AEM.00382-09

Xia, S., Duan, L., Song, Y., Li, J., Piceno, Y.M., Andersen, G.L., Alvarez-Cohen, L., Moreno-Andrade, I., Huang, C.-L., Hermanowicz, S.W., 2010. Bacterial community structure in geographically distributed biological wastewater treatment reactors. Environ. Sci. Technol. 44, 7391–7396. doi:10.1021/es101554m

Yang, J., 2014. Viral pathogens in clinical samples by use of a metagenomic approach, *in*: Nelson, K.E. (Ed.), Encyclopedia of Metagenomics. Springer New York, pp. 1–6.

Ye, L., Zhang, T., 2013. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Appl. Microbiol. Biotechnol. 97, 2681–2690. doi:10.1007/s00253-012-4082-4

Yu, K., Zhang, T., 2013. Construction of customized sub-databases from NCBI-nr database for rapid annotation of huge metagenomic datasets using a combined BLAST and MEGAN approach. PLoS One 8, e59831. doi:10.1371/journal.pone.0059831

Yu, K., Zhang, T., 2012. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7, e38183. doi:10.1371/journal.pone.0038183

Zhang, C., Qiu, S., Wang, Y., Qi, L., Hao, R., Liu, X., Shi, Y., Hu, X., An, D., Li, Z., Li, P., Wang, L., Cui, J., Wang, P., Huang, L., Klena, J.D., Song, H., 2014. Higher isolation of NDM-1 producing *Acinetobacter baumannii* from the sewage of the hospitals in Beijing. PLoS One 8, e64857. doi:10.1371/journal.pone.0064857

Zhang, T., Shao, M.-F., Ye, L., 2012. 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. ISME J. 6, 1137–1147. doi:10.1038/ismej.2011.188

Zhang, T., Zhang, X.-X., Ye, L., 2011. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One 6, e26041. doi:10.1371/journal.pone.0026041

Zhang, Y., Marrs, C.F., Simon, C., Xi, C., 2009. Wastewater treatment contributes to selective increase of antibiotic resistance among *Acinetobacter* spp. Sci. Total Environ. 407, 3702–3706. doi:10.1016/j.scitotenv.2009.02.013

Zheng, W., Zhang, Z., Liu, C., Qiao, Y., Zhou, D., Qu, J., An, H., Xiong, M., Zhu, Z., Zhao, X., 2015. Metagenomic sequencing reveals altered metabolic pathways in the oral microbiota of sailors during a long sea voyage. Sci. Rep. 5, 9131. doi:10.1038/srep09131

Zhou, C.E., Smith, J., Lam, M., Zemla, A., Dyer, M.D., Slezak, T., 2007. MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res. 35, D391–D394. doi:10.1093/nar/gkl791

2.8 Figures

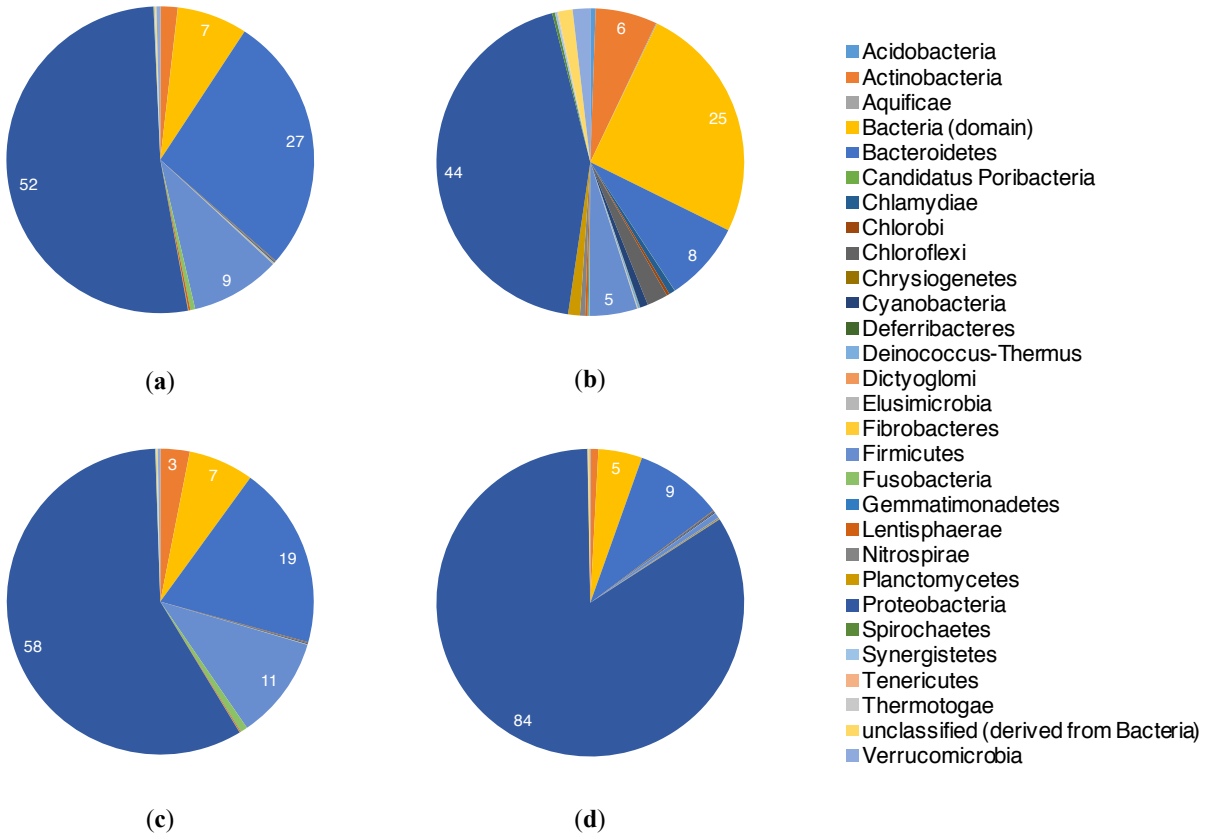


Figure 2.1: Occurrence and relative abundances of dominant phyla in Plant A (**a**) influent (**b**) effluent and Plant B (**c**) influent (**d**) effluent annotated by MG-RAST pipeline and classified with the lowest common ancestor algorithm. Relative abundance represents the number of reads affiliated with that phyla divided by the total reads assigned for the bacterial domain. Several phyla dominating the domain in the sample is indicated on the chart as percentages.

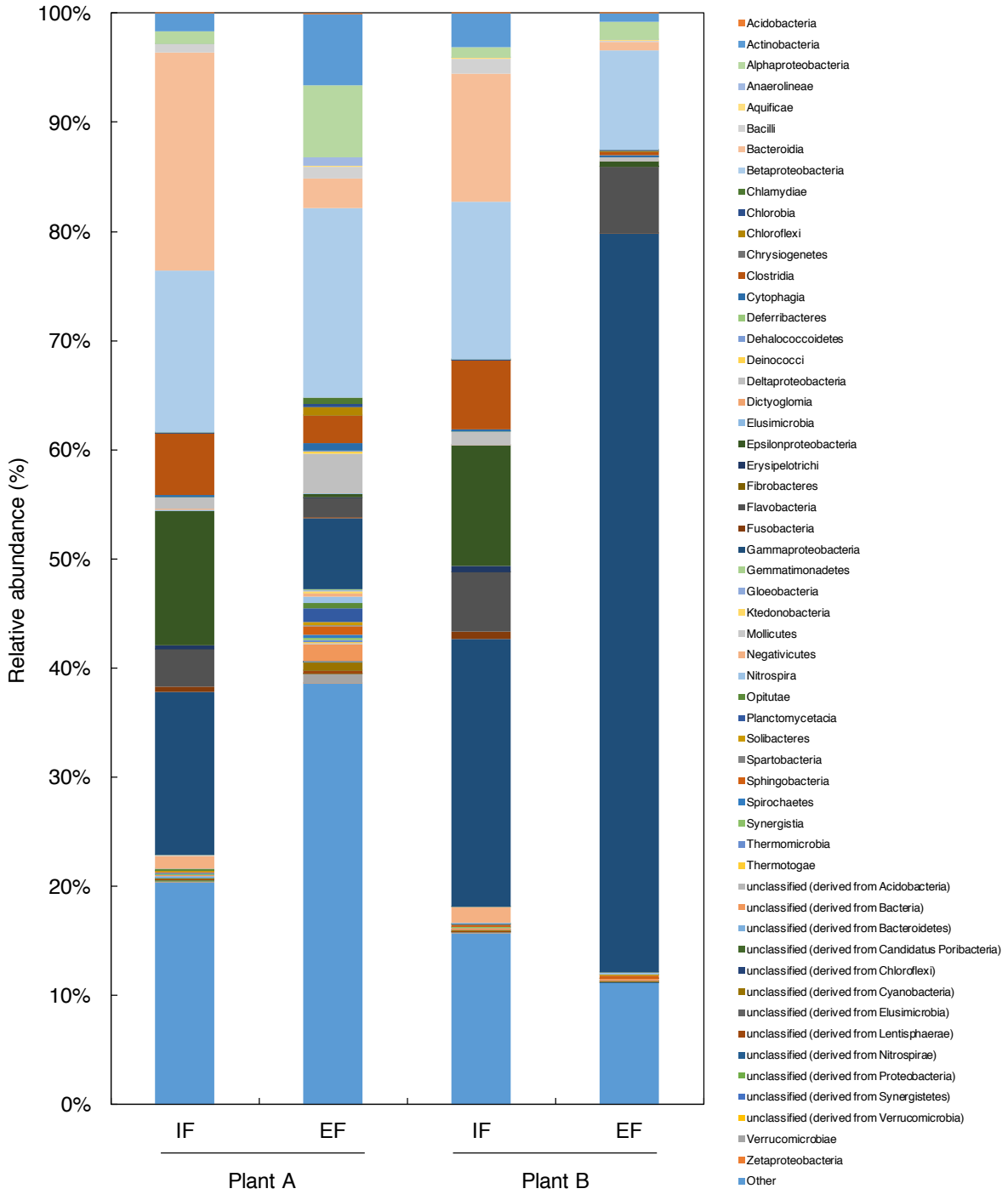
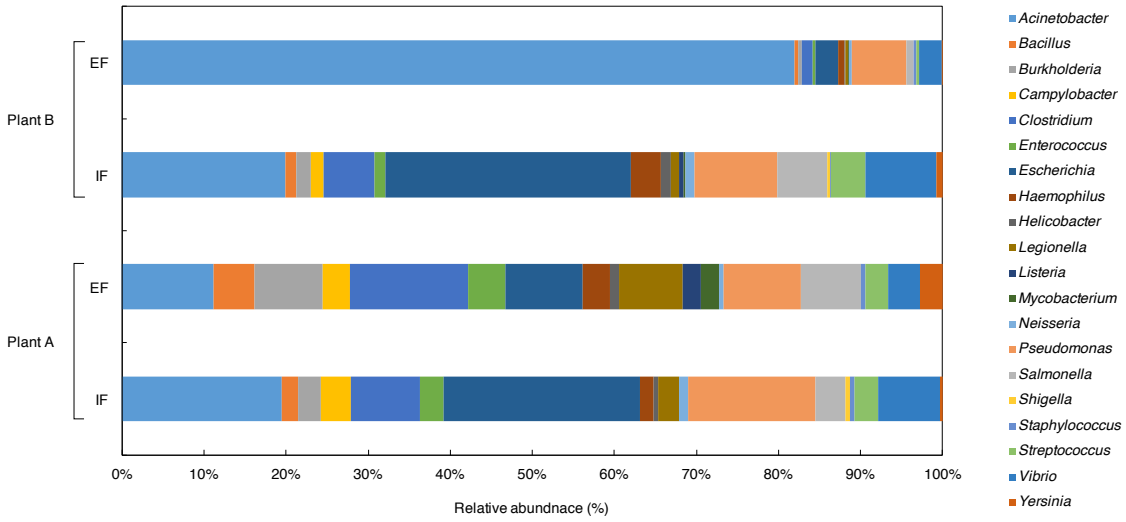
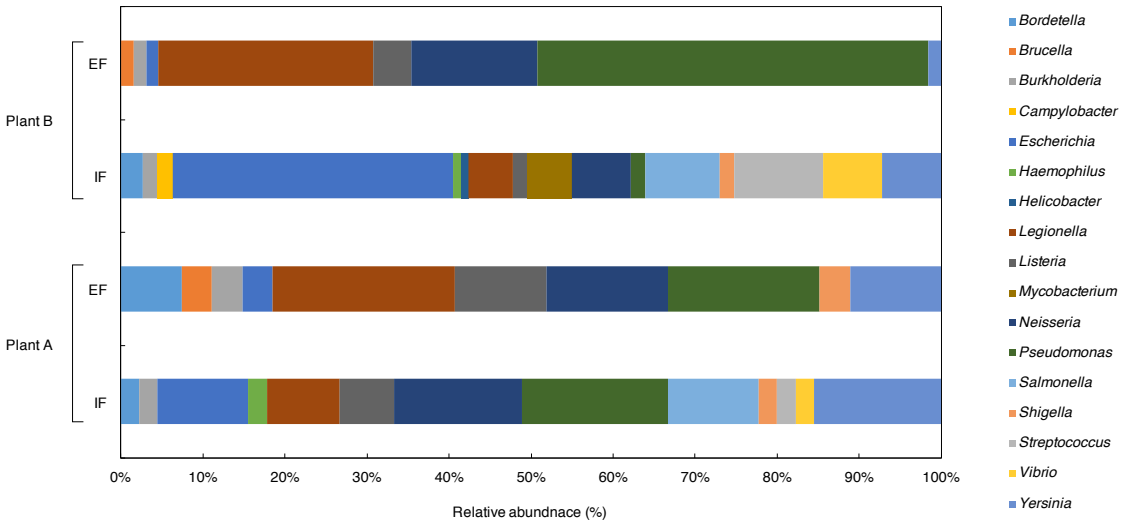


Figure 2.2: Occurrence and relative abundances of dominant classes observed in Plant A influent (IF). effluent (EF) and Plant B IF. EF annotated by MG-RAST pipeline and classified with the lowest common ancestor algorithm. Relative abundance represents the number of reads affiliated with that class divided by the total reads assigned for the bacterial domain. The "Other" category in the figure legend represents assignments to the class level rather than potential novel class (represented as "unclassified derived from").



(**a**)



(**b**)

Figure 2.3: Occurrence and relative abundances of genera containing potentially pathogenic species at the genus taxonomic level revealed by (**a**) annotation with the MG-RAST pipeline and classified with the lowest common ancestor algorithm and (**b**) annotation and classification with the human pathogenic bacteria virulence factor database. Relative abundance represents the number of reads affiliated with that genus divided by the total reads assigned in all genera.

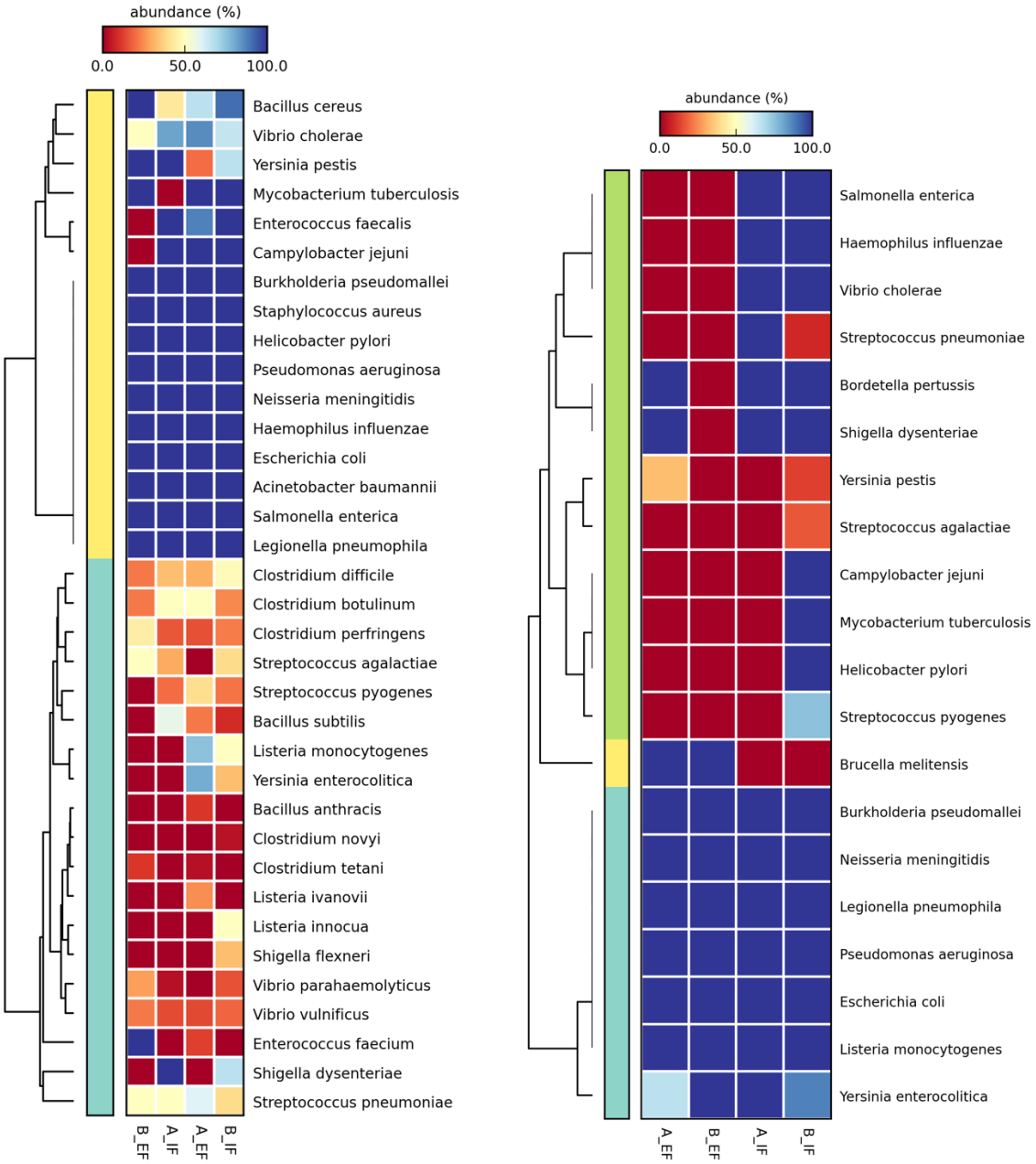


Figure 2.4: Heatmap depicting the distribution and relative abundance of potentially pathogenic species in Plant A influent (A_IF), effluent (A_EF), Plant B influent (B_IF) and effluent (B_EF) revealed by (**a**) annotation with the MG-RAST pipeline and classified with the lowest common ancestor algorithm and (**b**) annotation and classification with the human pathogenic bacteria virulence factor database. Relative abundance represents the number of reads affiliated with that genus divided by the total reads assigned in all genera.

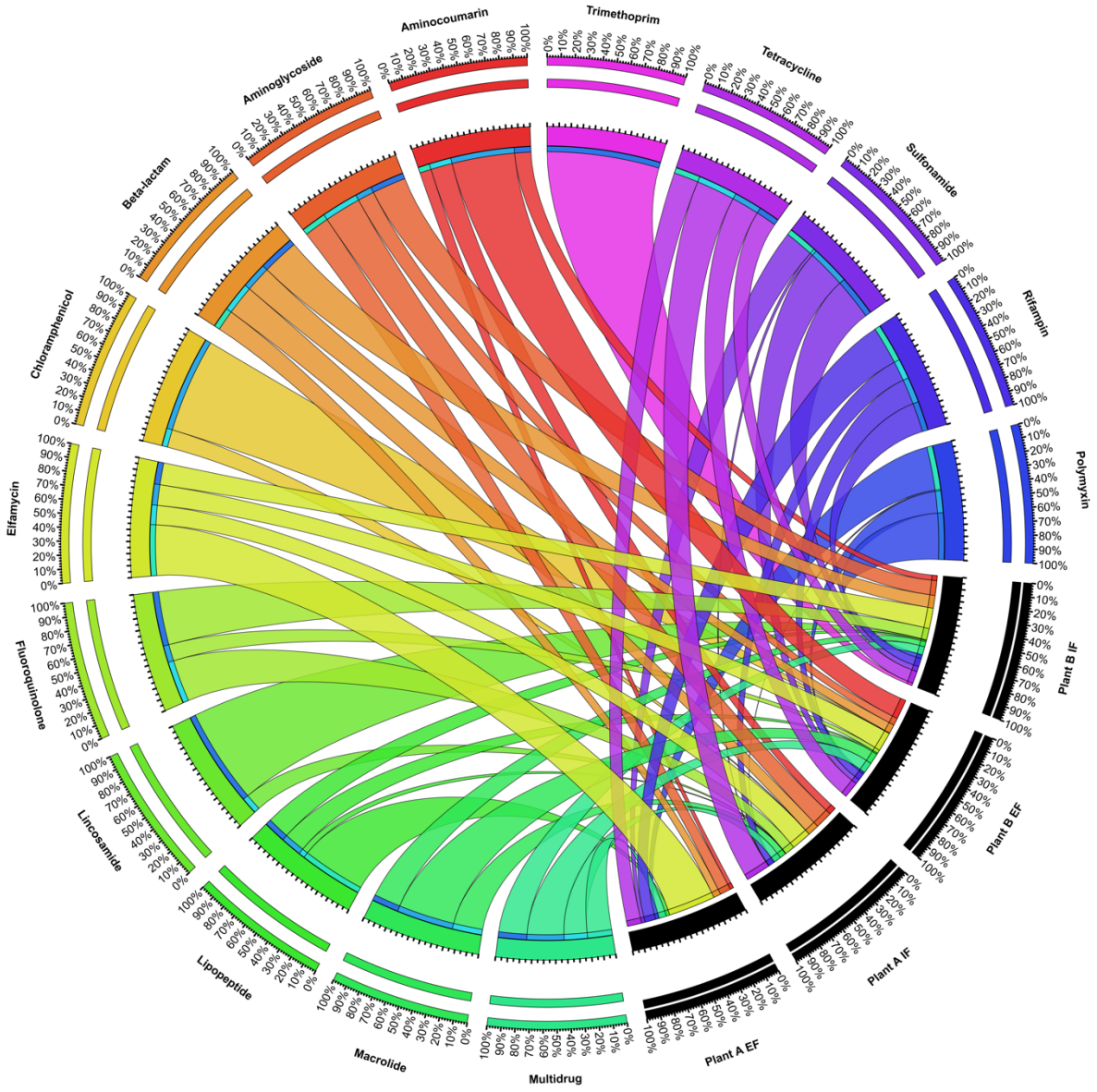


Figure 2.5: Distribution and relative abundance of antibiotic resistance genes (ARGs) types observed from Plant A influent (IF), effluent (EF) and Plant B IF and EF (visualized *via* Circos). Relative abundance values were calculated by dividing the number of annotated ARGs to the total number of ARGs detected in the metagenomes. Each antibiotic resistance type is represented by a specific ribbon colour and the width of the outer ring for each ribbon represents the percentage relative abundance of ARGs in the associated metagenomes. Plant A and Plant B associated metagenomes has been coloured black with remaining antibiotic types in a variety of colours.

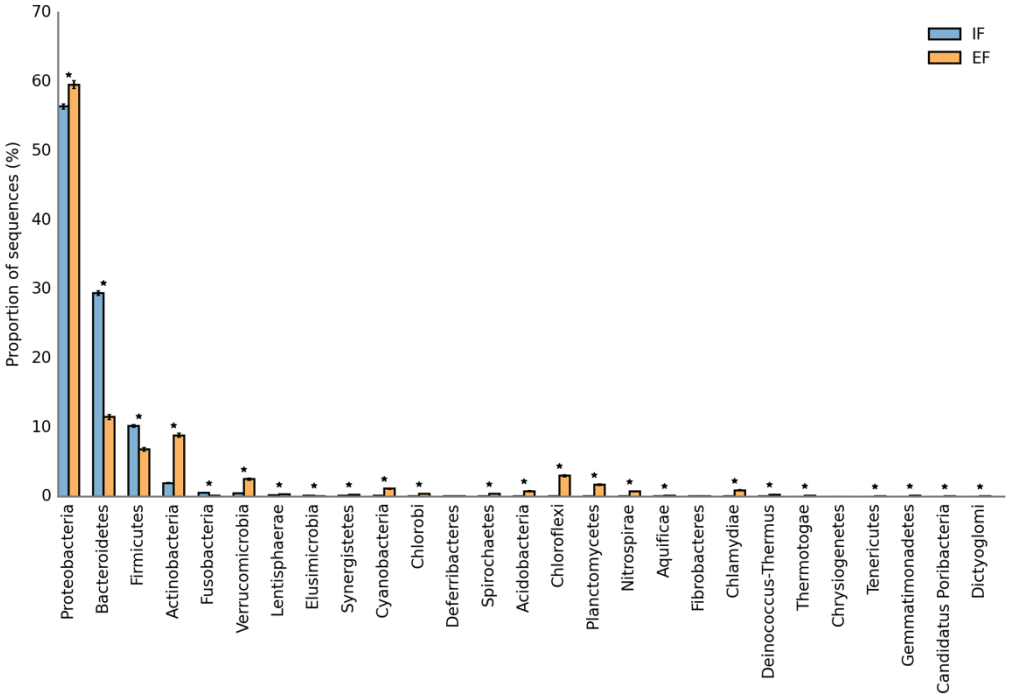
2.9 Supplementary material

Table S2.1: Characterization of the 454 pyrosequenced libraries from wastewater treatment plant (WWTP) influent and effluent microbial populations. Two different full-scale urban WWTPs located in Durban. South Africa were investigated in this study.

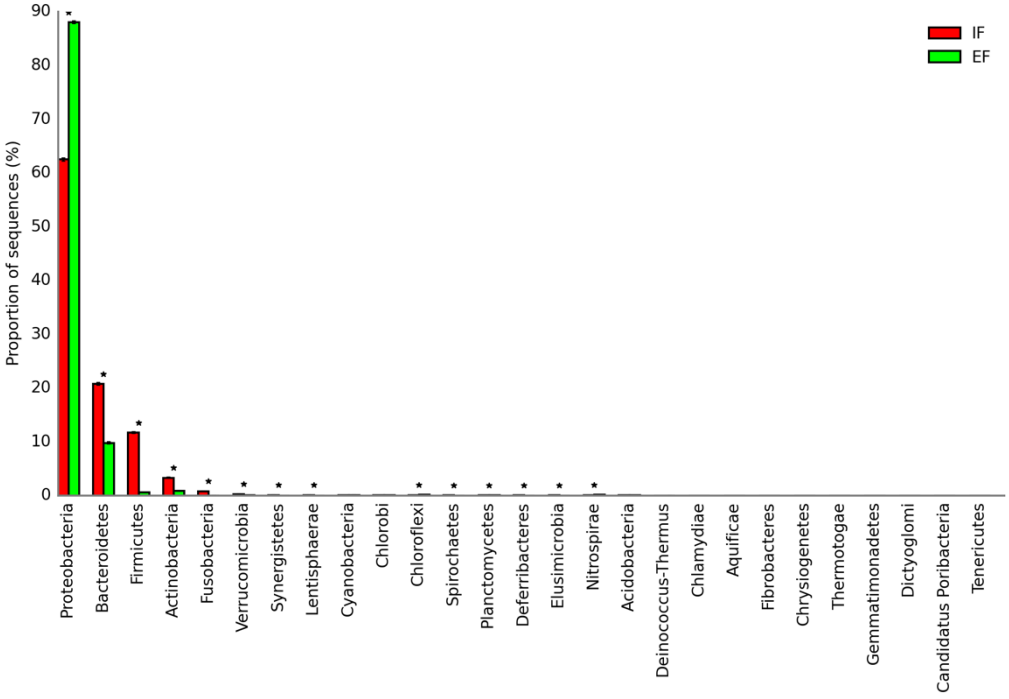
Characteristics	Plant A		Plant B	
	Influent	Effluent	Influent	Effluent
# raw reads	158 797	227 021	234 034	180 557
Total size (bp)	95 665 775	143 321 109	139 421 962	110 996 053
Average reads (bp)	602	631	596	615
# reads uploaded to MG-RAST	158 797	227 021	234 034	180 557
MG-RAST QC				
# reads before QC	158 797	227 021	234 034	180 557
Total size before QC (bp)	95 665 775	143 321 109	139 421 962	110 996 053
Average length before QC (bp)	602	631	596	615
# reads removed during ambiguous base filtering	14 981	11 511	21 023	30 777
# reads removed during dereplication	9 134	11 511	13 978	2 031
# reads after QC	134 682	194 388	199 033	147 749
Total size after QC (bp)	31 555 611	41 085 936	47 542 086	27 636 500
Average length after QC (bp)	234	211	239	187
MG-RAST Annotations				
# reads identified as RNA	22 235	32 299	32 975	23 065
# reads predicted ORFs	124 573	175 950	186 281	131 523
Database Annotations				
# reads assigned to VFDB	45	27	111	65
# reads assigned to CARD	99	31	175	187

Table S2.2: Relative abundance of the microbial populations from two full-scale wastewater treatment plants influent and effluent annotated by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Relative abundance is reported as percentages and represents the number of reads affiliated with that domain divided by the total reads assigned for all domain.

Domain	Plant A		Plant B	
	Influent	Effluent	Influent	Effluent
Bacteria	99.40	93.32	99.42	99.44
Archaea	0.14	1.47	0.13	0.17
Eukaryota	0.12	4.37	0.10	0.24
Viruses	0.30	0.47	0.31	0.11
Other sequences	<0.01	<0.01	<0.01	0.01
Unclassified sequences	0.04	0.37	0.04	0.03

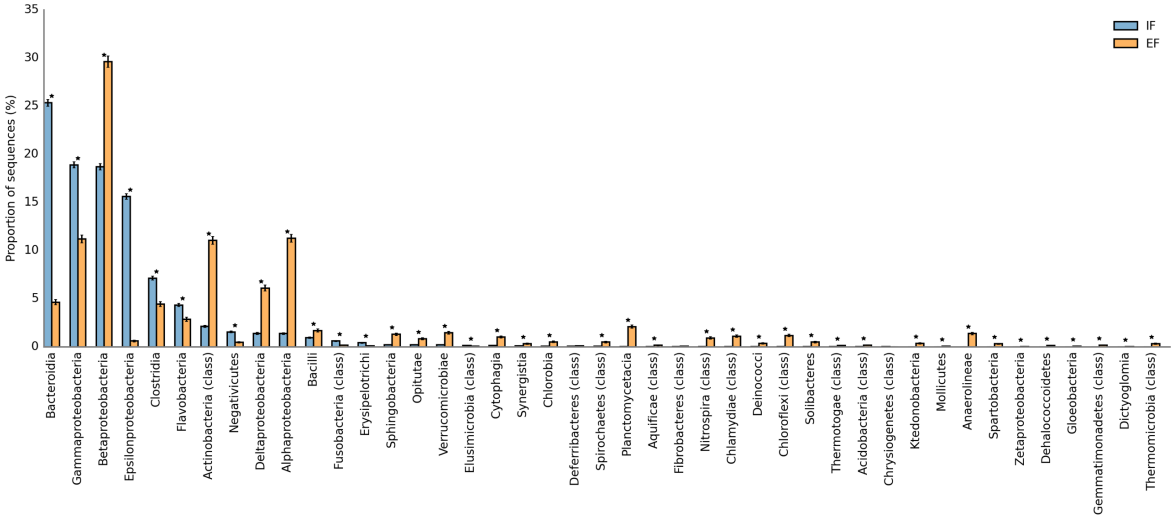


(**a**)

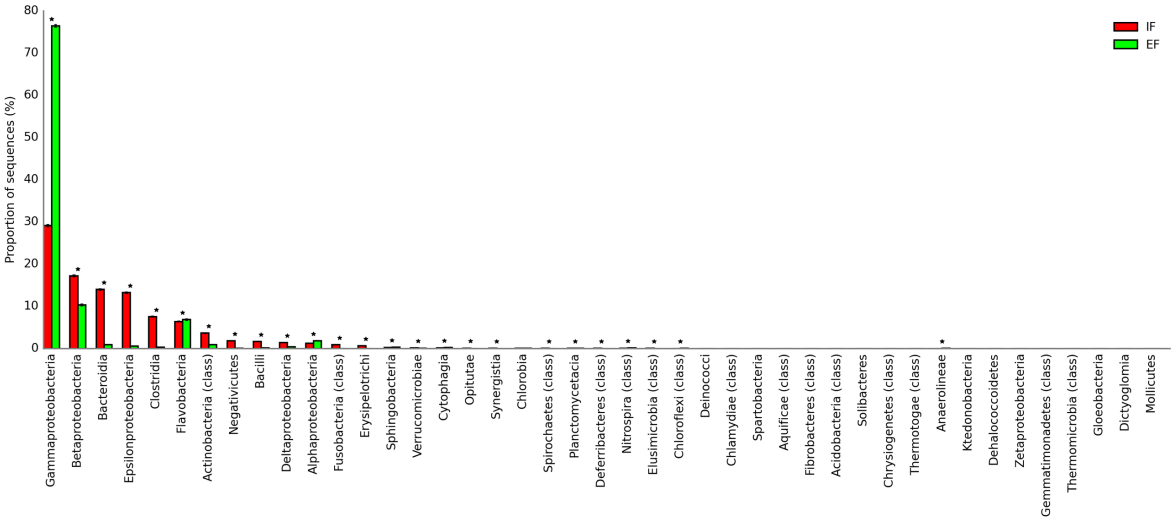


(**b**)

Figure S2.1: Comparative taxonomic profile of (**a**) Plant A influent (IF) and effluent (EF) and (**b**) Plant B IF and EF wastewater samples at phylum taxonomic level. Annotation was determined by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Asterisks on top of the bar graphs indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP.



(**a**)



(**b**)

Figure S2.2: Comparative taxonomic profile of (**a**) Plant A influent (IF) and effluent (EF) and (**b**) Plant B IF and EF wastewater samples at class taxonomic level. Annotation was determined by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Asterisks on top of the bar graphs indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP.

Table S2.3: Occurrence and relative abundance of potentially pathogenic bacterial species from two full-scale wastewater treatment plants influent (IF) and effluent (EF) annotated by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Relative abundance is reported as percentages and represents the number of reads affiliated with that species divided by the total reads assigned for all potentially pathogenic species.

Species	Plant A		Plant B	
	IF	EF	IF	EF
Acinetobacter baumannii	19.44	11.11	19.90	81.90
Pseudomonas aeruginosa	15.49	9.44	10.08	6.60
Escherichia coli	23.94	9.44	29.85	2.76
Vibrio cholerae	6.20	3.33	5.67	1.38
Salmonella enterica	3.66	7.22	6.05	0.92
Haemophilus influenzae	1.69	3.33	3.65	0.77
Vibrio parahaemolyticus	0.28	0	1.39	0.77
Clostridium perfringens	1.41	2.22	1.39	0.61
Vibrio vulnificus	1.13	0.56	1.64	0.61
Bacillus cereus	0.85	3.33	1.26	0.46
Burkholderia pseudomallei	2.82	8.33	1.76	0.46
Clostridium botulinum	4.23	7.22	1.51	0.31
Clostridium difficile	2.82	4.44	3.02	0.31
Enterococcus faecium	0	0.56	0	0.31
Legionella pneumophila	2.54	7.78	1.01	0.31
Neisseria meningitidis	1.13	0.56	1.13	0.31
Staphylococcus aureus	0.56	0.56	0.13	0.31
Clostridium tetani	0	0.56	0	0.15
Helicobacter pylori	0.56	1.11	1.26	0.15
Mycobacterium tuberculosis	0	2.22	0.25	0.15
Streptococcus agalactiae	0.85	0	1.64	0.15
Streptococcus pneumoniae	1.41	1.67	1.64	0.15
Yersinia pestis	0.28	0.56	0.50	0.15
Bacillus anthracis	0	0.56	0	0
Bacillus subtilis	1.13	1.11	0.13	0
Campylobacter jejuni	3.66	3.33	1.51	0
Clostridium novyi	0	0	0.25	0
Enterococcus faecalis	2.82	3.89	1.39	0
Listeria innocua	0	0	0.25	0
Listeria ivanovii	0	0.56	0	0
Listeria monocytogenes	0	1.67	0.25	0
Shigella dysenteriae	0.56	0	0.25	0
Shigella flexneri	0	0	0.13	0
Streptococcus pyogenes	0.56	1.11	0.88	0
Yersinia enterocolitica	0	2.22	0.25	0

Table S2.4: Occurrence and relative abundance of potentially pathogenic bacterial species from two full-scale wastewater treatment plants influent (IF) and effluent (EF) annotated and assigned with the virulence factor database. Relative abundance is reported as percentages and represents the number of reads affiliated with that species divided by the total reads assigned for all potentially pathogenic species.

Species	Strain	Plant A		Plant B	
		IF	EF	IF	EF
Pseudomonas aeruginosa	PAO1	17.78	18.52	1.8	47.69
Legionella pneumophila	subsp. pneumophila str. Philadelphia 1	8.89	22.22	5.41	26.15
Neisseria meningitidis	MC58	15.56	14.81	7.21	15.38
Listeria monocytogenes	EGD-e	6.67	11.11	1.8	4.62
Brucella melitensis	bv. 1 str. 16M	0	3.7	0	1.54
Burkholderia pseudomallei	K96243	2.22	3.7	1.8	1.54
Escherichia coli	CFT073	8.89	3.7	0.9	1.54
Yersinia enterocolitica	subsp. enterocolitica 8081	15.56	7.41	6.31	1.54
Bordetella pertussis	Tohama I	2.22	7.41	2.7	0
Campylobacter jejuni	subsp. jejuni NCTC 11168	0	0	1.8	0
Campylobacter jejuni	O157:H7 str. EDL933	2.22	0	31.53	0
Campylobacter jejuni	O45:K1:H7 str. S88	0	0	1.8	0
Haemophilus influenzae	Rd KW20	2.22	0	0.9	0
Helicobacter pylori	26695	0	0	0.9	0
Mycobacterium tuberculosis	H37Rv	0	0	5.41	0
Salmonella enterica	subsp. enterica serovar Typhi str. CT18	8.89	0	8.11	0
Salmonella enterica	subsp. enterica serovar Typhimurium str. LT2	2.22	0	0.9	0
Shigella dysenteriae	Sd197	2.22	3.7	1.8	0
Streptococcus agalactiae	2603V/R	0	0	1.8	0
Streptococcus pneumoniae	TIGR4	2.22	0	0.9	0
Streptococcus pyogenes	M1 GAS	0	0	8.11	0
Vibrio cholerae	O1 biovar El Tor str. N16961	2.22	0	7.21	0
Yersinia pestis	CO92	0	3.7	0.9	0

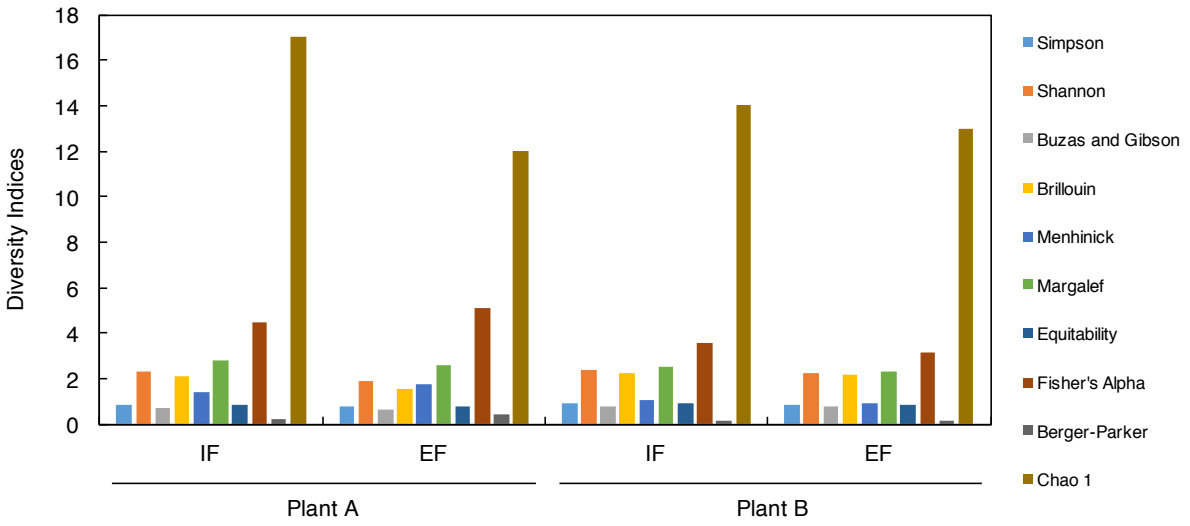


Figure S2.3: Diversity indices (Simpson, Shannon, Buzas and Gibson, Brillouin, Menhinick, Margalef, Equitability, Fisher's alpha, Berger-Parker and Chao 1) of the abundance of ARGs types determined for Plant A WWTP influent (IF), effluent (EF), Plant B WWTP IF and EF.

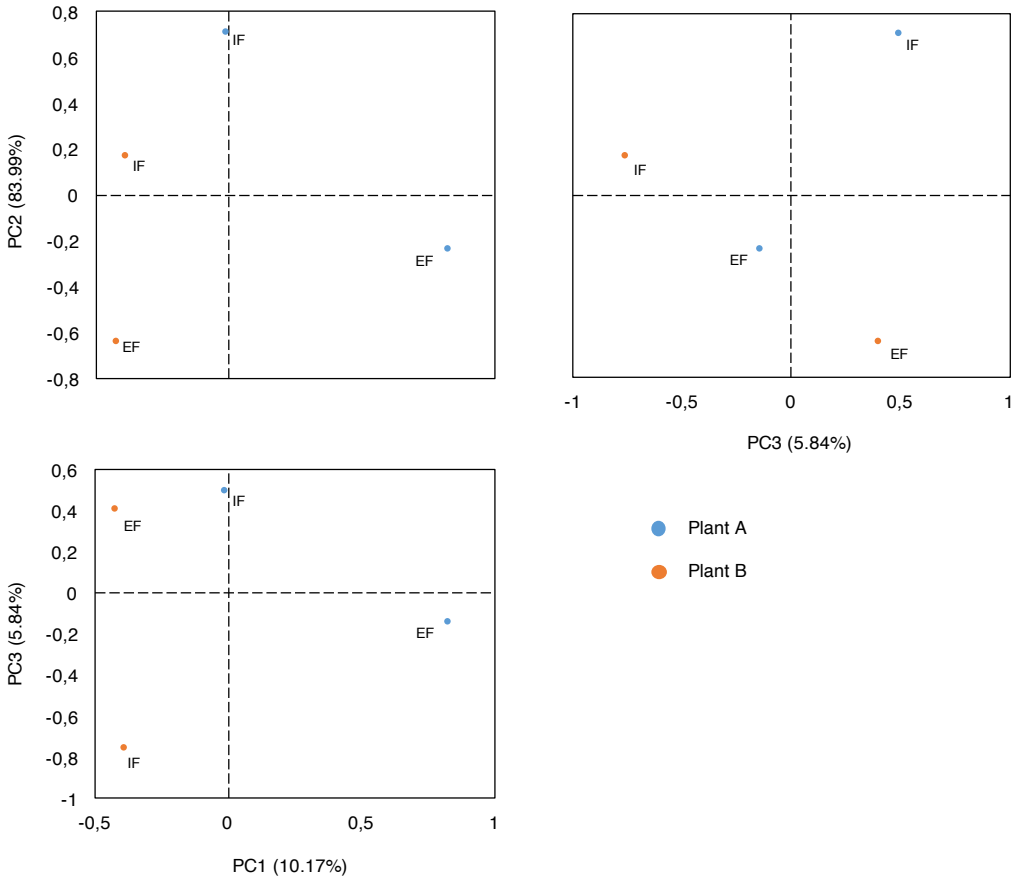


Figure S2.4: Principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity coefficients of different ARGs types determined for Plant A WWTP influent (IF), effluent (EF), Plant B WWTP IF and EF.

Chapter Three

Manuscript formatted and submitted to *Water Research*

3 Chlorination effects on bacterial diversity, functional potential and antibiotic resistome during urban wastewater treatment: a metagenomic insight

3.1 Abstract

The most common method of disinfection used by wastewater treatment plants (WWTPs) aimed at removing microbial pathogens from the treated effluent is chlorination. However, some studies have reported on the selective pressure exerted by the process for antibiotic resistance. In this study, we investigated the effects of chlorination on bacterial diversity and structure, functional potential and antibiotic resistome of wastewater communities using whole-genome shotgun pyrosequencing combined with metagenomic analysis in two full-scale WWTPs. Overall, chlorination significantly influenced the microbial diversity and structure in both WWTPs, creating a shift between Beta-, Gamma- and Alpha-proteobacteria in pre- and post-chlorination communities, suggesting the ability of the members of these classes to resist disinfection treatment. This was further supported by the increase in specific protective mechanisms, such as glutathione metabolism and oxidative stress subsystems in post-chlorination communities at both WWTPs. Overall decrease in antibiotic resistome was observed following disinfection suggesting the effectiveness of this treatment stage. However, chlorination seems to promote the selection of resistance to the elfamycin and tetracycline classes of antibiotics at both WWTPs. The results from this study shed more light on the influence of chlorination disinfection on the bacterial community dynamics, functional potential and antibiotic resistome during chlorination treatment in WWTPs.

Keywords: Antibiotic resistance; Chlorination; Metagenomics; Microbial communities; Pyrosequencing; Wastewater

3.2 Introduction

Wastewater treatment plants (WWTPs) are an essential urban infrastructure. The most common method of wastewater disinfection internationally adopted for the removal of pathogens in wastewater effluent is chlorination. Wastewater disinfection is essential for ensuring compliance to water quality standards before treated wastewater effluents are discharged from WWTPs into receiving rivers, streams or the ocean (Blatchley et al., 2007). Although disinfection by chlorination has been shown to be effective in decreasing the load of various bacteria, viruses and protozoa, chlorination does not completely eradicate the growth of microorganisms (Gomez-Alvarez et al., 2012). Conversely, previous findings have suggested that urban WWTPs act as major reservoirs for residual antibiotics (AB), antibiotic resistant genes (ARGs) and antibiotic resistance bacteria (ARB) and are responsible for their dissemination into the environment (Kümmerer, 2009; Lupo et al., 2012; Rizzo et al., 2013). The concern is the potential in which the dissemination of wastewater effluents into the environment would contribute to the generation and selection of ARB capable of infecting human and animals (Courvalin, 2008; Da Silva et al., 2006; Davies and Davies, 2010; Figueira et al., 2011; Olaniran et al., 2012; Wellington et al., 2013), leading to animal and public health risks (Ram et al., 2008; WHO, 2014).

Previous studies have reported on the diverse microbial species and communities found in wastewater effluents discharged from WWTPs using a variety of culture- or molecular-dependent techniques, each with its own advantages and limitations (Hong et al., 2010; Olaniran et al., 2015, 2012; Revetta et al., 2011; Rizzo et al., 2013). Specifically, several studies have shown the contribution of chlorination in the spread of ARGs (Huang et al., 2011; Karumathil et al., 2014) and enrichment for ARB (Armstrong et al., 1982; Karumathil et al.,

2014; Murray et al., 1984). However, most of these studies have been limited in scope as the use of culture-dependent methods to comprehensively characterize microbial communities remains a challenge. Firstly, it is a time consuming task with most environmental organisms remaining unculturable *in vitro* with current available technologies (Schmieder and Edwards, 2012). Secondly, the inert PCR bias associated with microbial diversity fingerprinting and PCR based-clone library methods do not allow for a comprehensive and unbiased view of the complex microbial communities present in wastewater samples (Yadav et al., 2014).

Recently, high-throughput sequencing combined with metagenomic analysis of a given sample has been considered a promising approach for the assessment of complex microbial communities (Kristiansson et al., 2011; Wang et al., 2013; Zhang et al., 2011). This approach has been applied for the assessment of various environmental scenarios. These include the analysis of soil (Howe et al., 2014) and river (Amos et al., 2014) microbiota, communities found in activated sludge from urban WWTPs (Ju et al., 2014), communities in sediments from an oil spill (Mason et al., 2014), faecal microbiota (Ilmberger et al., 2014), microbial diversity in aquatic settings (Doxey et al., 2014) as well as clinical settings such as the human gut microbial profiling project (Qin et al., 2010), sputum analysis of patients with cystic fibrosis (Lim et al., 2014), viral pathogens detection (Yang, 2014) and many others. In determining the complex genetic composition and diversity present in WWTP samples with a relatively unbiased view (Gomez-Alvarez et al., 2009), a metagenomic insight is a favourable approach. Furthermore, recent studies have indicated the importance of understanding bacterial community shifts as it is said to influence the shaping of the microbial community resistome (Forsberg et al., 2014; Su et al., 2015).

In this study, high-throughput pyrosequencing of metagenomic samples combined with bioinformatic analysis were used for the assessment of the diversity and composition of the bacterial communities present in wastewater samples. Furthermore, the community functional potential and antibiotic resistome was also determined. Results from this study may help to extend our knowledge on the complex microbial communities found in urban WWTPs and the effects of chlorination on the microbial communities during this important disinfection treatment stage.

3.3 Materials and Methods

3.3.1 WWTP description and sample collections

In this study, pre-chlorinated and post-chlorinated wastewater samples were collected from two full-scale urban WWTPs in the city of Durban, South Africa, hereon designated as Plant A and Plant B. Plant A has a capacity of 70 megaliters/day with an operational capacity of 96% and uses the activated sludge and diffused air liquid technologies with gravity thickening, anaerobic digestion and belt press dewatering sludge technologies. Plant B has a capacity of 25 megaliters/day with an operational capacity of 76% and uses the activated sludge liquid technology with anaerobic digestion and belt press dewatering sludge technologies (Department of Water and Sanitation, 2015). Furthermore, Plant A receives a mixture of nearby domestic and industrial wastewater, whilst Plant B receives a mixture of domestic, industrial and hospital wastewaters. Four samples were collected in 5 L plastic bottles pre-sterilized with 70% (v/v) ethanol and rinsed with 4 L of the sample at the various sampling sites prior to

collection. Upon collection, the samples were transported on ice back to the laboratories within 3 h and stored at 4°C prior to DNA extraction which took place within 24 h.

3.3.2 Ethics statement

No special permits were required for this study. Permission for collection of pre-chlorinated and post-chlorinated samples from both WWTPs was granted by the authorities of the respective WWTPs (Durban, South Africa).

3.3.3 Total DNA extraction and shotgun pyrosequencing

Total DNA of the bacterial populations were extracted from the samples using PowerWater™ DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. This kit is specifically designed for isolating bacterial DNA from environmental water samples and includes inhibitor removal technology aimed at removing humic acid and other organic matter commonly found in environmental samples that can interfere with downstream analyses. The resulting purity and concentration of the DNA preparation was determined using the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, USA) before samples were sent to the National Genomics Platform (Technology Innovation Agency, South Africa) for sequencing.

High-throughput whole-genome shotgun pyrosequencing was conducted by the use of 454 GS FLX System (Roche, USA) with the GS FLX Titanium Rapid Library Preparation Kit (Roche, USA). The samples were barcoded by the use of GS FLX Titanium Rapid Library MID

Adaptors Kit (Roche, USA) in order to enable multiplex sequencing. DNA of 500 µg was used for library construction. Briefly, DNA was nebulized using nitrogen gas and purified using MinElute PCR Purification Kit (QIAGEN, Germany) according to manufacturer's instructions. Fragment end repair and attachment of adaptors to the samples was carried out according to standard protocol (Roche, USA). Thereafter, the quality of the libraries was assessed by the use of 2000 Bioanalyzer (Agilent Technologies, USA) using the High Sensitivity DNA Analysis Kit (Agilent Technologies, USA). Thereafter, relative fluorescence of the DNA libraries was measured on GloMax-Multi Microplate Multimode Reader (Promega, USA) using 6-fluorescein amidite (6-FAM) standards and 6-FAM-labeled Multiplex Identifiers (MIDs). The results were uploaded onto the Rapid Library Quantitation Calculator (Roche, USA) and rapid libraries were made with a final concentration of 1×10^7 molecules/µl. Emulsion PCR was carried out with the use of GS FLX Titanium MV emPCR Kit (Lib-L) (Roche, USA) according to manufacturer's instructions. Subsequently, four DNA copies per bead were enriched and used for the main sequencing run with *Escherichia coli* beads as a positive control (Roche, USA).

3.3.4 Data availability

All individual sequence reads have been deposited at the NCBI Short Read Archive (SRA) under the accessions no. SRR3629047 and SRR3629048 for Plant A pre- and post-chlorination datasets, respectively, and SRR3629053 and SRR3629057 for Plant B pre- and post-chlorination datasets, respectively.

3.3.5 Bioinformatic analysis of pyrosequencing datasets

Raw data files from sequencing were de-multiplexed by removing the barcoded sequence and any secondary adapter sequences. For initial quality control, raw reads were evaluated by CLC Genomics Workbench v.7.5 (CLC Bio, QIAGEN, Germany) quality control pipeline. The quality of all reads was within the acceptable standards from the platform (see Table S3.1 in the supplemental material). Therefore, no reads were trimmed, filtered or discarded before implementation of the MG-RAST v.3.6 pipeline for further quality control, prediction and annotation (Meyer et al., 2008; Wilke et al., 2015). Ambiguous base filtering by removing sequences with >5 ambiguous base pairs (bp) and length filtering by removing sequences with a length of >2 standard deviations from the mean were applied as quality control filters using SolexaQA (Cox et al., 2010) in the MG-RAST pipeline.

3.3.5.1 Taxonomic classifications

For taxonomic classification, an initial BLAT search against reduced RNA database (90% identity clustered version of SILVA database) was performed and the rRNA-similar reads were then clustered at 97% identity with the longest read as the cluster representative. Thereafter, BLAT search of the cluster representative was performed against the M5rna database (see Table S3.1 in the supplemental material). M5rna is a ribosomal database with integration of SILVA, Greengenes and RDP databases (Meyer et al., 2008; Wilke et al., 2015). Taxonomic abundance was analyzed using the Lowest Common Ancestor (LCA) algorithm used in MEGAN (Huson et al., 2007) and implemented in MG-RAST with a maximum e-value cut-off of 1×10^{-5}, minimum identity of 60% and a minimum alignment length of 15, which is measured in amino acids for protein databases and bp for RNA databases (Zheng et al., 2015).

This algorithm assigns each read to the LCA from the set of matching taxa when BLASTx is applied. For example, if a given read had sequence similarity to 3 different families within the same order, the read is assigned at the order level rather than assigning the read to a specific family. Hence, the LCA algorithm has been reported to have lower rates of false positive assignments than the best hit classification algorithm implemented in the MG-RAST pipeline. However, this would result in a higher number of unspecific assignments or no hits in some cases (Huson et al., 2007).

3.3.5.2 Functional annotations

Prior to functional annotation, reads within the datasets were screened for artificially replicated sequences and those identified sequences were removed using the dereplication tool (Gomez-Alvarez et al., 2009) implemented in the MG-RAST pipeline (see Table S1 in the supplemental material). Thereafter, putative ORFs were identified using FragGeneScan, an *ab-initio* prokaryotic gene calling algorithm using the Hidden Markov Model (Rho et al., 2010), and their corresponding protein sequences were searched with BLAST against the M5nr protein database in the MG-RAST server. M5nr is a non-redundant protein database with integration of many public sequences databases including GenBank, SEED, IMG, UniProt, KEGG and eggNOGs (Meyer et al., 2008; Wilke et al., 2015). For functional assignments, analysis was performed using Hierarchical Classification approach against KEGG orthology (KO) and SEED subsystems (SS) with maximum e-value cut-off, minimum identity and a minimum alignment length as described above.

3.3.6 Comparative metagenomic analyses

To determine the statistical differences in taxonomic and functional distribution between metagenomic datasets, statistical analysis of metagenomic profiles (STAMP) v.2.1.3 software package was used (Parks and Beiko, 2010). Statistical significance of differences between samples (q values) was calculated on the basis of two-sided Fisher's exact test using Storey's false discovery rate (FDR). Because p values were not uniformly distributed using Storey's FDR, Benjamin-Hochberg FDR was applied for corrections during functional annotations. The confidence intervals were determined by Newcombe-Wilson's method. Only features with a q value of <0.05 were considered significant in this study.

Principal Coordinate Analysis (PCoA) of the Bray-Curtis dissimilarity coefficients based on the normalized abundance of functional assignments under the SS was used to identify the relationships between functional community structures of wastewater metagenomes in this study and 48 publically available metagenomes covering a wide variety of habitats (see Table S2 in the supplemental material). All publically available metagenomes utilized the same sequencing method as described for this study with the exception of Artic freshwater datasets which utilized Sanger sequencing. Annotation and functional assignment of the publically available metagenomes was performed on the MG-RAST server to avoid potential bias from different analysis pipelines (Jeffries et al., 2011; Smith et al., 2012). Functional assignments were performed using Hierarchical Classification approach against SS with maximum e-value cut-off, minimum identity and a minimum alignment length as described above. Prior to PCoA, SS assignments of each metagenomic datasets were normalized against the total number of hits in the respective database, log transformed and implemented in MG-RAST to avoid variations.

3.3.7 Detection of antibiotic resistome

To identify ARGs in the metagenomes in order to establish the antibiotic resistome profiles, the Comprehensive Antibiotic Resistance Database (CARD) was used (McArthur et al., 2013). CARD database was preferred over the other ARGs databases, such as Antibiotic Resistance Genes Online (Scaria et al., 2005), the microbial database of protein toxins, virulence factors, and antibiotic resistance genes (MvirDB) (Zhou et al., 2007), and Antibiotic Resistance Genes Database (Liu and Pop, 2009), because these databases are neither exhaustive nor regularly updated (Gupta et al., 2014). BLASTx against the database was carried out using DIAMOND tool v.0.7.11 (Buchfink et al., 2014) with a e-value cut-off of 1×10^{-5}. A read with its best BLAST hit was deemed ARG-like if the hit had a sequence similarity of above 90% over an alignment of at least 25 amino acids (Chao et al., 2013; Wang et al., 2013). The classification of ARG-like sequences was performed using the structured database of CARD (McArthur et al., 2013) and in-house written scripts. One-way analysis of variance (ANOVA) was carried out to assess the variations between the different antibiotic resistome of the pre- and post-chlorination communities using IBM SPSS v.22 (Armonk, New York: IBM Corp.), and $p<0.05$ was considered statistically significant. Principal coordinate analysis was performed to evaluate the difference between the antibiotic resistome of the samples based on the Bray-Curtis dissimilarity coefficients of the relative abundance of ARGs. Various diversity indices for the antibiotic resistome were determined using the 'vegan' package (Shen and Fulthorpe, 2015) from the program R (R Development Core Team 2007).

3.4 Results

Whole-genome shotgun metagenomics in combination with pyrosequencing were used to gain an insight into the effect of chlorination on the taxonomic diversity and composition, functional potential and antibiotic resistome of microbial communities in wastewater samples of two full-scale WWTPs.

In total, approximately 914,987 raw reads with an average length of 539 bp corresponding to approximately 479 million bp were generated for this study. Specifically, raw sequences generated were 237,014 and 227,021 reads for pre- and post-chlorinated samples in Plant A, respectively, and 270,395 and 180,557 reads for pre- and post-chlorinated samples in Plant B, respectively. Upon ambiguous base filtering and length filtering, 9.3% of the reads in both pre- and post-chlorination datasets for Plant A, and 19.7% and 18.2% of the reads for pre- and post-chlorinated samples in Plant B, respectively, was excluded from further analyses. Prior to functional annotation, approximately 5% of the reads for pre- and post-chlorinated samples in Plant A, and 5% of pre-chlorination and 1% of post-chlorination for Plant B metagenomes were identified as artificially replicated sequences and removed (see Table S3.1 in the supplemental material).

3.4.1 Microbial composition and diversity

Taxonomic classification of both predicted rRNA genes sequences and predicted proteins in all metagenomic datasets were conducted based on all the available annotation source databases in MG-RAST. Domain distributions in the four samples, showed the expected

dominance of bacteria across all metagenomes occupying 97.81% and 93.32% for pre- and post-chlorinated samples from Plant A, respectively, and 99.01% and 99.44% for pre- and post-chlorinated samples from Plant B, respectively. Also, low numbers of the Oligohymenophorea class of eukaryotic members were found in abundance, after the bacterial domain, accounting for 0.35% and 2.8% of the annotated sequences in pre- and post-chlorinated samples of Plant A, respectively. Whilst, low numbers of the Streptophyta phylum of eukaryotic members were found in abundance, after the bacterial domain, accounting for 0.14% and 0.04% of the annotated sequences in pre- and post-chlorinated samples of Plant B, respectively. The remaining sequences were distributed amongst archeal and viral domains in very low abundance (see Table S3.3 in the supplemental material).

At the phyla taxonomic level, bacterial communities were diverse with representatives of 27 phyla in pre- and post-chlorinated samples in both Plant A and Plant B. Statistical analysis using STAMP indicated that no correlation (r^2=0.873) was found at this taxonomic level between bacterial community diversity in pre- and post-chlorinated samples in Plant A, whilst high positive correlations (r^2=0.989) were seen between pre- and post-chlorinated samples in Plant B at this taxonomic level (see Figure S3.1 in the supplemental material). In all wastewater samples, Proteobacteria was the dominant phylum accounting for 43.68% to 83.81% of annotated reads in effluent of both WWTPs. This was followed by Bacteroidetes, Actinobacteria and Firmicutes phyla (Figure 3.1). Disinfection treatment in Plant A was observed to show a 21.8% reduction in the Proteobacteria phylum from pre- to post-chlorination, whilst Plant B showed a 14.3% increase in the reads classified. Additionally, significant differences (q<0.05) were observed at this level amongst Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes between pre- and post-chlorinated wastewaters

from both WWTPs (see Figure S3.2 in the supplemental material). Notably, Actinobacteria, Firmicutes and Chloroflexi significantly increased after disinfection treatment in Plant A. However, significant decrease was observed for these phyla after disinfection in Plant B.

At the class taxonomic level, Betaproteobacteria was the most abundant class accounting for 26.68% of the total bacterial sequences, followed by Gammaproteobacteria (23.46%), Flavobacteria (8.34%) and Alphaproteobacteria (3.87%) in pre-chlorinated wastewater of Plant A. In post-chlorinated wastewater of Plant A, Betaproteobacteria was the most abundant class accounting for 17.37% of total bacterial sequences, followed by Alphaproteobacteria (6.61%), Gammaproteobacteria (6.57%) and Actinobacteria (6.49%) (Figure 3.2). Additionally, significant differences (q<0.05) were observed for Betaproteobacteria, Gammaproteobacteria, Flavobacteria and Alphaproteobacteria amongst others at the class level (see Figure S3.3 in the supplemental material). Gammaproteobacteria was the most abundant class in pre-chlorinated wastewater of Plant B, accounting for 38.20% of total bacterial sequences, followed by Betaproteobacteria (17.11%), Flavobacteria (6.22%) and Actinobacteria (3.40%). In post-chlorinated wastewater of Plant B, Gammaproteobacteria was still the most abundant class accounting for 67.69% of total bacterial sequences, followed by Betaproteobacteria (9.16%), Flavobacteria (6.07%) and Alphaproteobacteria (1.63%) (Figure 3.2). However, significant differences (q<0.05) were observed for Gammaproteobacteria, Betaproteobacteria, Actinobacteria and other phyla at this level (see Figure S3.4 in the supplemental material). It should be noted that a 3.6-fold decrease in members of the Gammaproteobacteria and a 1.7-fold and 3-fold increase in members of the Alphaproteobacteria and Actinobacteria class was observed upon disinfection treatment in Plant A, whilst a 2-fold increase of the members of the

Gammaproteobacteria class was observed following disinfection treatment in Plant B (Figure 3.2).

3.4.2 Functional potential of microbial communities

Predicted protein features were annotated using MG-RAST with SS and KO. In all cases, to avoid bias from lower quality spurious hits, an e-value threshold of 1×10^{-5} was applied when the results were exported or further processed. Overall, annotation of the metagenomic sequences against the MG-RAST M5nr database with hierarchical classification revealed a total of 77,212 and 42,663 reads were assigned to the various level 1 SS for the wastewater communities in Plant A (for pre- and post- chlorination, respectively). Whilst, 92,346 and 71,314 reads were assigned to the various level 1 SS wastewater communities in Plant B (for pre- and post- chlorination, respectively). For KO assignments, a total of 32,724 and 19,038 reads were assigned to class 1 KO wastewater communities in Plant A (pre- and post-chlorination, respectively), whilst a total of 40,346 and 31,009 reads were assigned to class 1 KO wastewater communities in Plant B (pre-and post-chlorination, respectively).

The dominant functional level 1 subsystems for bacterial communities in Plant A (pre- and post-chlorination) were “clustering-based subsystems”, followed by “protein metabolism” and “carbohydrates” in decreasing relative abundance. For Plant B pre-chlorination, the dominant subsystems among the bacterial communities were “clustering-based subsystems”, “carbohydrates”, “amino acids and derivatives” and “protein metabolism” in decreasing relative abundance. However, shifts were observed, post-chlorination, with “clustering-based subsystems”, “amino acids and derivatives”, “miscellaneous” and “protein metabolism” in

decreasing relative abundance as the dominant subsystems (Figure 3.3 and see Table S3.4 and S3.5 in the supplemental material).

Statistical analysis of the overall level 1 subsystem of communities in pre- and post-chlorinated wastewater of Plant A and Plant B showed no correlation in their functional potential (r^2=0.817 and r^2=0.874, respectively) (Figure 3.4a and 3.4b). Comparison of the level 1 subsystems between pre- and post-chlorinated wastewaters in Plant A indicated that 15 out of the 28 subsystems were significantly different (q<0.05) (Figure 3.4c). Specifically, significant increases were observed in "clustering-based subsystems", "carbohydrates", "protein metabolism", "DNA metabolism", "phages, prophages, transposable elements, plasmids" and "cell division and cell cycle" subsystems in post-chlorinated wastewaters of Plant A. Conversely, significant decreases were observed in "amino acids and derivatives", "cofactors, vitamins, prosthetic groups, pigments", "fatty acids, lipids, and isoprenoids", "virulence, disease and defence", metabolism of aromatic compounds", "motility and chemotaxis", nitrogen metabolism", iron acquisition and metabolism" and "sulphur metabolism". Comparison of the level 1 SEED subsystems in pre- and post-chlorination communities of Plant B indicated that 14 out of the 28 subsystems were significantly different (q<0.05) (Figure 4d). Specifically, significant increase was observed in "cofactors, vitamins, prosthetic groups, pigments", "RNA metabolism", "fatty acids, lipids, and isoprenoids", "nucleosides and nucleotides", "metabolism of aromatic compounds", "regulation and cell signalling", "iron acquisition and metabolism", "sulphur metabolism" and "secondary metabolism" subsystems. Conversely, "carbohydrates", "protein metabolism", "DNA metabolism", "nitrogen metabolism" and "motility and chemotaxis" subsystems decreased significantly.

Detailed analysis of the disinfectant resistance mechanisms of the bacterial communities include the oxidative stress response functions assigned in SS and the glutathione synthesis pathways assigned in KO. Specifically, the OxyR system, SoxRS system, RpoS regulated genes and exopolysaccharide (EPS) synthesis pathways assigned in SS were investigated.

In bacterial communities of pre- and post-chlorinated wastewater of both Plants, 11 out of 21 enzymes involved in the glutathione synthesis pathways was annotated and assigned to KEGG pathway (Figure 3.5a and 3.6a). All three of the enzymes for bacterial glutathione synthesis (viz. glutathione synthase, glutathione reductase and glutathionylspermidine synthase) were detected at considerable abundances (Figure 3.5a and 3.6a). Comparing the abundance of the genes encoding the enzymes of the pathway in pre- and post-chlorinated wastewater of Plant A, higher abundance of leucyl aminopeptidase, dipeptidase D and 5-oxoprolinase was observed among communities post-chlorination. Contrarily, higher abundance of glutathione reductase, leucyl aminopeptidase, glutamate-cysteine ligase and 5-oxoprolinase was observed for Plant B. For the OxyR system, alkyl hydroperoxide reductase protein C (*ahpC*), glutaredoxin 1 reductase (*grxA*) and thioredoxin reductase (*trxB*) were found in greater abundance post-chlorination in Plant A (Figure 3.5b), whilst alkyl hydroperoxide reductase protein F (*ahpF*), hydrogen peroxide-inducible regulator (*oxyR*), peroxidase/catalase (*katG*) and *trxB* were found in greater abundance post-chlorination in Plant B (Figure 3.6b). For the SoxRS system, aconitate hydratase/aconitase A (*acnA*), glucose-6-phosphate 1-dehydrogenase/G6PD (*zwf*), magnesium superoxide dismutase (*sodA*) and redox-sensitive transcriptional regulator (*soxR*) was found in higher abundance post-chlorination in Plant A (Figure 3.5c), whilst only fumarate hydratase/fumarase C (*fumC*), *sodA* and *soxR* were found in higher abundance post-chlorination in Plant B (Figure 3.6c). For the RpoS system, all three enzymes, Cu-Zn

superoxide dismutase (*sodC*), exo-DNase III (*xthA*) and RNA polymerase sigma factor (*rpoS*) were detected in greater abundance post-chlorination in Plant A (Figure 3.5d), whilst all three enzymes were observed in low abundance post-chlorination in Plant B (Figure 3.6d). Genes involved in exopolysaccharide (EPS) synthesis pathways, primarily belonging to cell wall and capsule SS was determined for both all wastewater samples. Among these, genes for EPS biosynthesis and dTDP-rhamnose synthesis was found in greatest abundance in post-chlorinated wastewaters of Plant A and Plant B, respectively (Figure 3.5e and 3.6e).

Comparison of the functional potential of bacterial communities of all wastewater samples to 6 different environmental habitats by principle coordinate analysis (PCoA) revealed that different ecosystems clustered differently (Figure 3.7). For example, human oral, human gut, ocean and soil metagenomes cluster into individual groups. Conversely, Arctic freshwater and hydrocarbon wastewater samples showed variations in metagenomes although sampled from the same ecosystem. Wastewater samples from this study were observed to ordinate around ocean, soil and Artic freshwater metagenomes.

3.4.3 Antibiotic resistome abundance and diversity

BLASTx against the CARD protein database showed that 94 reads (pre-chlorination) and 31 reads (post-chlorination) were assigned to a total of 15 classes of ARGs in Plant A (Figure 3.8a). For Plant B wastewaters, BLASTx against the CARD database showed that 169 and 187 reads were assigned to pre- and post-chlorinated wastewaters, respectively, with a total of 14 classes of resistance genes (Figure 3.8b). The total abundance of the detected ARGs showed no significant differences (p>0.05) between the pre- and post-chlorinated wastewaters at both

WWTPs. However, Shannon and Simpson's diversity indices calculations consistently indicated that disinfection by chlorination reduced the diversity of the ARGs in treated effluents of both WWTPs (see Figure S3.5 in the supplemental material). Additionally, PCoA of the antibiotic resistome patterns revealed shifts after disinfection of the wastewaters in both WWTPs, where each of the antibiotic resistome profile clustered separately from each other (see Figure S3.6 in the supplemental material). This was further observed with additional calculations of diversity indices (see Table S3.5 in the supplemental material).

The elfamycin resistance genes was the dominant class in Plant A wastewater communities (pre- and post-chlorination). This was followed by tetracycline, aminoglycoside, aminocoumarin and macrolide classes in decreasing relative abundance in pre-chlorinated wastewater. Post-chlorination, elfamycin resistance gene was followed by tetracycline, aminoglycoside, rifampin, beta-lactam classes and multidrug efflux pump genes in decreasing relative abundance. In pre-chlorinated wastewater in Plant B, elfamycin resistance gene was also the dominant class of ARGs, followed by tetracycline, macrolide, beta-lactam, aminoglycoside classes and multidrug efflux pump genes in decreasing relative abundance. Post-chlorination, bacterial communities showed aminocoumarin as the dominant class, followed by elfamycin, tetracycline, macrolide, multidrug efflux pump genes and beta-lactam classes in decreasing relative abundance.

Under the elfamycin ARG class, variants of the elongation factor Tu were the dominant type of ARG in pre- and post-chlorinated wastewater bacterial communities of both WWTPs. Relative abundance of 29.79% and 41.94% of the total ARGs was observed, pre- and post-

chlorination respectively in Plant A, while a relative abundance of 17.16% and 17.11% of the total ARGs was observed in pre- and post-chlorinated wastewaters, respectively in Plant B, (Figure 3.8c). The *msrE* and *tet39* genes were found to be in high abundance, pre-chlorination in Plant A. However, post-chlorination, *APH(3")-Ib* and *rpoB* genes conferring resistance to rifampicin was found to be in higher abundance. Similarly, *msrE* and *tet39* genes were also found in high abundance in pre-chlorinated wastewater of Plant B, while, *tet39* and *adeJ* genes were in higher abundance in post-chlorination communities.

3.5 Discussion

3.5.1 Microbial composition and diversity

Metagenomic analysis consistently indicated that chlorine disinfection altered the microbial community diversity of the wastewater from the two different WWTPs investigated in this study, although Proteobacteria was observed to persist after the disinfection treatment used by both WWTPs. This result is consistent with studies using PCR-based experiments targeting the 16S rRNA gene to examine the microbial communities of samples from WWTPs (Miura et al., 2007; Silva et al., 2010a, 2010b). The result is also consistent with a study which reported Proteobacteria as the dominant persisting phylum during drinking water disinfection by chlorination (Chao et al., 2013; Poitelon et al., 2009; Shi et al., 2013). Furthermore, investigations on the community structure of biofilms in drinking water treatment plants during disinfection with a metagenomic approach also showed that the dominant microbes were related to Proteobacteria phylum (Chao et al., 2015; Schmieder and Edwards, 2012). Dominance by this phylum could be due to the fact that this phylum comprises one of the most phylogenetically and metabolically diverse group in the Bacteria domain (Ettema and

Andersson, 2009). Besides dominance by Proteobacteria, dominance by Bacteroidetes, Actinobacteria and Firmicutes phyla in the bacterial communities were seen in this study. This is also consistent with previously reported studies investigating community structure of microbes in untreated wastewaters (Shanks et al., 2013), swine WWTP (Da Silva et al., 2015), anaerobic reactor digesting activated sludge from WWTP (Guo et al., 2015), tannery WWTP (Wang et al., 2013) and activated sludge of WWTP in Hong Kong (Yu and Zhang, 2012). It has been suggested that the core human microbial signature is made up of Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria, with shifts in the ratio of Firmicutes and Bacteroidetes correlating with obesity (Buffie and Pamer, 2013). As both WWTPs in this study receives domestic wastewaters and hospital effluents (in the case of Plant B), it is not surprising that the bacterial community is dominated by members of these phyla due to the source of influent received by the WWTPs. It is noteworthy that for this study we did not remove free associated DNA or dead cells before DNA isolation and therefore the sequences identified could be associated with live and/or dead microbial populations. Whilst, urban WWTPs present microbes with a relatively harsh environment, especially during treatment stages, there have been reports which indicate survival of pathogenic ARB which survive all treatment stages (Odjadjare and Olaniran, 2015; Olaniran et al., 2015, 2012).

Shifts in the microbial community structure was observed at the class level in both WWTPs upon disinfection. In Plant A, members of the Beta- and Gamma-proteobacteria dominated the community before disinfection treatment, however, a shift to Beta- and Alpha-proteobacteria was observed, post-chlorination. This suggests that Beta- and Alpha-proteobacteria may tolerate chlorination disinfection during wastewater treatment. Shifts to Alpha-proteobacteria was also observed following disinfection of drinking water (Chao et al., 2013), moving bed

biofilm reactor systems treating municipal sewage (Biswas and Turner, 2012) and investigations on the bacterial communities in different sections of a municipal WWTP (Ye and Zhang, 2013). Conversely, in Plant B, Gamma- and Beta-proteobacteria dominated the community before and after disinfection of the wastewater with no shift observed. This result is consistent with a report which indicated the shift towards decreased relative abundance of Betaproteobacteria, and increased relative abundance of Firmicutes, Alphaproteobacteria, Gammaproteobacteria in the microbial ecology of a hospital's hot water system after supplemental monochloramine disinfection implementation (Baron et al., 2014). Therefore, overall results observed in this study suggests that selected members from Proteobacterial classes, Beta-, Gamma- and Alpha-proteobacteria, may tolerate chlorination better than the remaining classes under the Proteobacteria phylum.

3.5.2 Functional potential of microbial communities

Whole genome shotgun sequencing of metagenomes has the capacity to fully sequence the majority of available genomes within an environmental sample (or community). Therefore, besides community biodiversity profiles, functional potential of microbial communities can be explored. As with the diverse bacterial structure observed in this study, a wide metabolic diversity was also present in the wastewater communities of both WWTPs.

As expected, genes assigned to the metabolism of carbohydrates, protein metabolism, amino acids and proteins were observed to be in relatively high abundance as they are related to the housekeeping functions of all living organisms. However, clustering-based subsystems (containing functions such as proteosomes, ribosomes, and recombination-related clusters

(Delmont et al., 2012)), and miscellaneous subsystems (containing genes associated with iron-sulphur cluster assembly, common prokaryotic-plant genes and Niacine-Choline transport and metabolism) were also found to be in relatively high abundance. These results were observed in previous studies investigating soil (Delmont et al., 2012), marine (Gilbert et al., 2008), activated sludge (Dinsdale et al., 2008) and freshwater (Dinsdale et al., 2008) communities. Comparison of the overall functional potential between other ecological settings revealed considerable functional differences between different ecosystems. Though, different ecosystems show distinct functional characteristics, wastewaters from this study shared similar functional characteristics with ocean and soil sources. These observations are expected as WWTPs present with an environment closer to soil and ocean environments.

Significant increases in several higher level subsystems and the shift observed in the metabolic potential upon disinfection with no correlations between pre- and post-chlorinated wastewater suggests that these functions could be important in providing the resistance to chlorination treatment. Specifically, direct increase in bacterial resistance to chlorine compounds have been reported to include protective genes involved in glutathione metabolism (Chesney et al., 1996). Several genes involved in the bacterial pathway of glutathione metabolism (Chao et al., 2013) were observed in this study. The widely accepted and proposed theory of glutathione metabolism in eukaryotes is by the transfer of these genes from bacteria via the ancestral predecessor of mitochondria during evolution (Fahey et al., 1984). This suggests that the modern relatives of the ancestral predecessor of mitochondria, members of the Alphaproteobacteria class, should house glutathione biosynthesis genes, and might explain the increase in the abundance over other bacterial classes observed, post-chlorination in Plant A. Moreover, because the nutrient levels present in treated wastewaters prior to disinfection

treatment should be at low levels, starvation of bacterial species is expected to stimulate glutathione synthesis and subsequently enhance chlorine resistance (Saby et al., 1999). Additionally, genes involved in EPS synthesis, although more related to biofilm production, could also provide protective functions (Ryu and Beuchat, 2005) and were observed in the wastewater communities at both WWTPs. Furthermore, genes involved in Glutathione metabolism has been said to be indirectly involved in the regulation of several oxidation resistant systems, such as OxyR, SoxR and SOS systems (Saby et al., 1999).

Specifically, activation of the OxyR system can induce the transcription of various antioxidant genes, such as *trxB*, *ahpCF*, *grxA*, *katG* and non-specific DNA-binding protein (*dps*) (Sund et al., 2008). The GrxA protein acts as a repair mechanism for the oxidized form of glutathione in many Gram-negative bacteria (Gallardo-Madueño et al., 1998; Müller et al., 1995; Smirnova et al., 2001), and the reduction of disulphide bonds in proteins damaged by oxidative radicals is carried out by TrxB (Gómez-Pastor et al., 2010; Kuhns et al., 2015; Russel and Model, 1988). These genes were found in high abundance post-chlorination in both WWTPs, suggesting that OxyR systems could be an important resistance mechanism for chlorine. This is supported by studies which have suggested the key role of the OxyR system in resistance (Chao et al., 2015; Gray et al., 2013; Wang et al., 2009). *acnA* and *fumC* genes are involved in the TCA cycle whereby the TCA cycle maintain cellular redox balance by reduction reactions (Zheng et al., 1999), whilst activation of the *zwf* gene would produce NADPH/NADH in response to severe oxidative stress (Shimizu, 2013; Sund et al., 2008). Genes involved in responses to the SoxR system, including *acnA* and *zwf* were found in high abundance post-chlorination in Plant A, suggesting that the microbial communities in these wastewaters potentially followed the NADPH/NADH synthesis and TCA cycle. Conversely, Plant B post-chlorination communities

saw high abundance in only *fumC* and potentially followed only the TCA cycle. The presence of these genes could explain the poor efficiency of chlorination for inactivation and the reduction of pathogens abundance during disinfection process.

3.5.3 Antibiotic resistome abundance and diversity

Chlorination was found to effectively decrease the relative abundance of ARGs in Plant A, which agrees with a recent study (Zhang et al., 2015). Surprisingly, increase in the relative abundance of ARGs in Plant B was observed post-chlorination. This result was also observed in a study investigating drinking water (Chao et al., 2013). Overall, a decrease in the diversity of ARGs upon chlorination in both WWTPs was observed. This suggests that there is a relationship between decreased diversity and decrease in total ARGs abundance in Plant A. Conversely, for Plant B, chlorination decreased the diversity, but increased the total abundance of the ARGs in the bacterial communities. Several studies investigating the ARGs using molecular methods have also reported the same trend observed for Plant B (Mao et al., 2015; Shi et al., 2013; Xi et al., 2009). Overall, the various antibiotic classes observed post-chlorination in this study was also reported in a previous study investigating ARGs from environments impacted by livestock and municipal waste (Agga et al., 2015).

Unexpectedly, elfamycin genes were found to be the most abundant ARGs in both WWTPs. The elfamycin family of antibiotics inhibits protein synthesis of Gram-negative bacteria by an interaction with elongation factor Tu (Hall et al., 1989; Vogeley et al., 2001; Wolf et al., 1974). This family is said to be subdivided into three types on the basis of their structure. Specifically, majority of the elfamycin ARGs in the wastewater communities were associated with the

variants of the elongation factor Tu. Though elfamycin are not commonly used in treatment of human diseases due to the very narrow spectrum against human pathogens (Miele et al., 1994), it is used as a growth promoting agent in animals owing to its excellent activities (Brötz et al., 2011; Maiese et al., 1989). Therefore, high abundance of the ARGs suggests that the major source of wastewaters received by both WWTPs could be from nearby animal feed industrial wastewaters. Furthermore, it suggests that chlorination during wastewater treatment does not effectively decrease the relative abundance of this family of antibiotics and this warrants further investigations. This could also partially explain the high abundance of *tet39* genes observed post-chlorination as tetracycline is not effectively biodegradable and commonly used for livestock purposes in South Africa (Eagar et al., 2012). Furthermore, studies suggests that sources of *tet39* gene could result from clinical (Agersø and Guardabassi, 2005) or animal sources (Agersø and Petersen, 2007) and are common amongst resistant *Acinetobacter* spp (Hamidian et al., 2016), which is a cause for concern as chlorination does not seem to be effective in the removal of these ARGs.

It should be noted that the diversity and abundance of the antibiotic resistome observed by CARD alignment were different from those derived from the MG-RAST analysis of the metagenome. Possible reasons for the divergence may be the difference in the reference database used as well as the contribution on the alignment algorithm. MG-RAST relies on the use of BLAT algorithm for similarity searches, which is less sensitive than the BLAST algorithm (Yu and Zhang, 2013). Furthermore, CARD database is a highly curated and comprehensive database (McArthur et al., 2013) compared to the incomplete SEED "resistance to antibiotics and toxic compounds" subsystem on ARGs (http://theseed.uchicago.edu/FIG/subsys.cgi).

3.6 Conclusion

Detection of taxonomic profiles, functional potential and antibiotic resistome across the wastewater samples tested in this study further supports the use of the pyrosequencing approach as a preliminary investigation before deep-sequencing by Illumina technology is used to identify trends across diverse and complex environments. Taxonomic profiles were in general agreement with previous descriptions of WWTPs, in which Proteobacteria were the predominant members of the wastewater communities. Furthermore, the functional potential of these communities were also in general agreement with previously reported studies. However, our study indicated that selected members of the Proteobacterial class showed tolerance towards chlorination, and were detected in higher abundance after disinfection. Resistance to chlorine was supported by the detection of Glutathione metabolism and related oxidative stress response systems genes in the communities suggesting that they play a major as protective functions during disinfection of wastewaters. Results from this study further indicate the diversity of ARGs in wastewaters, and suggests for the first time, the dominance of the elfamycin family of ARG in urban WWTPs. These findings strongly support the notion that urban WWTPs are potential reservoirs for AR, ARGs and ARB selection. Further investigation on the effects of the effluent discharge from these WWTPs on the receiving surface water bodies is a subject on on-going investigation in our laboratory to allow for a more comprehensive understanding of the microbial ecology of WWTPs and the environment, and could be useful for the safeguard of public and environmental health.

3.7 References

Agersø, Y., Guardabassi, L., 2005. Identification of Tet 39, a novel class of tetracycline resistance determinant in *Acinetobacter* spp. of environmental and clinical origin. J. Antimicrob. Chemother. 55, 566–9. doi:10.1093/jac/dki051

Agersø, Y., Petersen, A., 2007. The tetracycline resistance determinant Tet 39 and the sulphonamide resistance gene sulII are common among resistant *Acinetobacter* spp. isolated from integrated fish farms in Thailand. J. Antimicrob. Chemother. 59, 23–7.

Agga, G.E., Arthur, T.M., Durso, L.M., Harhay, D.M., Schmidt, J.W., 2015. Antimicrobial-resistant bacterial populations and antimicrobial resistance genes obtained from environments impacted by livestock and municipal waste. PLoS One 10, e0132586. doi:10.1371/journal.pone.0132586

Amos, G.C.A., Zhang, L., Hawkey, P.M., Gaze, W.H., Wellington, E.M., 2014. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet. Microbiol. 171, 441–447. doi:10.1016/j.vetmic.2014.02.017

Armstrong, J.L., Calomiris, J.J., Seidler, R.J., 1982. Selection of antibiotic-resistant standard plate count bacteria during water treatment. Appl. Environ. Microbiol. 44, 308–316.

Baron, J.L., Vikram, A., Duda, S., Stout, J.E., Bibby, K., 2014. Shift in the microbial ecology of a hospital hot water system following the introduction of an on-site monochloramine disinfection system. PLoS One 9, e102679. doi:10.1371/journal.pone.0102679

Biswas, K., Turner, S.J., 2012. Microbial community composition and dynamics of moving bed biofilm reactor systems treating municipal sewage. Appl. Environ. Microbiol. 78, 855–64. doi:10.1128/AEM.06570-11

Blatchley, E.R., Gong, W.-L., Alleman, J.E., Rose, J.B., Huffman, D.E., Otaki, M., Lisle, J.T., 2007. Effects of wastewater disinfection on waterborne bacteria and viruses. Water Environ. Res. 79, 81–92.

Brötz, E., Kulik, A., Vikineswary, S., Lim, C.T., Tan, G.Y., Zinecker, H., Imhoff, J. F., Paululat, T., Fiedler, H.P., 2011. Phenelfamycins G and H, new elfamycin-type antibiotics produced by *Streptomyces albospinus* Acta 3619. J. Antibiot. 64, 257–266.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods advance on. doi:10.1038/nmeth.3176

Buffie, C.G., Pamer, E.G., 2013. Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol. 13, 790–801. doi:10.1038/nri3535

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X.-X., Wu, W.-M., Zhang, T., 2013. Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci. Rep. 3. doi:10.1038/srep03550

Chao, Y., Mao, Y., Wang, Z., Zhang, T., 2015. Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing. Sci. Rep. 5, 10044. doi:10.1038/srep10044

Chesney, J.A., Eaton, J.W., Mahoney, J.R., 1996. Bacterial glutathione: a sacrificial defense against chlorine compounds. J. Bacteriol. 178, 2131–5.

Courvalin, P., 2008. Predictable and unpredictable evolution of antibiotic resistance. J. Intern. Med. 264, 4–16. doi:10.1111/j.1365-2796.2008.01940.x

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485. doi:10.1186/1471-2105-11-485

Da Silva, M.F., Tiago, I., Veríssimo, A., Boaventura, R.A.R., Nunes, O.C., Manaia, C.M., 2006. Antibiotic resistance of enterococci and related bacteria in an urban wastewater treatment plant. FEMS Microbiol. Ecol. 55, 322–329. doi:10.1111/j.1574-6941.2005.00032.x

Da Silva, M.L.B., Cantão, M.E., Mezzari, M.P., Ma, J., Nossa, C.W., 2015. Assessment of bacterial and archaeal community structure in Swine wastewater treatment processes. Microb. Ecol. 70, 77–87. doi:10.1007/s00248-014-0537-8

Davies, J., Davies, D., 2010. Origins and evolution of antibiotic resistance. Microbiol. Mol. Biol. Rev. 74, 417–433. doi:10.1128/MMBR.00016-10

Delmont, T.O., Prestat, E., Keegan, K.P., Faubladier, M., Robe, P., Clark, I.M., Pelletier, E., Hirsch, P.R., Meyer, F., Gilbert, J.A., Le Paslier, D., Simonet, P., Vogel, T.M., 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. ISME J. 6, 1677–87. doi:10.1038/ismej.2011.197

Department of Water and Sanitation, 2015. Green Drop Progress Report. Pretoria, South Africa.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., Olson, R., Paul, J., Brito, B.R., Ruan, Y., Swan, B.K., Stevens, R., Valentine, D.L., Thurber, R.V., Wegley, L., White, B.A., Rohwer, F., 2008. Functional metagenomic profiling of nine biomes. Nature 452, 629–32. doi:10.1038/nature06810

Doxey, A.C., Kurtz, D.A., Lynch, M.D.J., Sauder, L.A., Neufeld, J.D., 2014. Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. ISME J. doi:10.1038/ismej.2014.142

Eagar, H., Swan, G., van Vuuren, M., 2012. A survey of antimicrobial usage in animals in South Africa with specific reference to food animals. J. S. Afr. Vet. Assoc. 83, 16. doi:10.4102/jsava.v83i1.16

Ettema, T.J.G., Andersson, S.G.E., 2009. The alpha-proteobacteria: the Darwin finches of the bacterial world. Biol. Lett. 5, 429–32. doi:10.1098/rsbl.2008.0793

Fahey, R.C., Newton, G.L., Arrick, B., Overdank-Bogart, T., Aley, S.B., 1984. *Entamoeba histolytica*: a eukaryote without glutathione metabolism. Science 224, 70–2.

Figueira, V., Vaz-Moreira, I., Silva, M., Manaia, C.M., 2011. Diversity and antibiotic resistance of *Aeromonas* spp. in drinking and waste water treatment plants. Water Res. 45, 5599–5611. doi:10.1016/j.watres.2011.08.021

Forsberg, K.J., Patel, S., Gibson, M.K., Lauber, C.L., Knight, R., Fierer, N., Dantas, G., 2014. Bacterial phylogeny structures soil resistomes across habitats. Nature 509, 612–616. doi:10.1038/nature13377

Gallardo-Madueño, R., Leal, J.F., Dorado, G., Holmgren, A., López-Barea, J., Pueyo, C., 1998. *In vivo* transcription of *nrdAB* operon and of *grxA* and *fpg* genes is triggered in *Escherichia coli* lacking both thioredoxin and glutaredoxin 1 or thioredoxin and glutathione, respectively. J. Biol. Chem. 273, 18382–8.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., Joint, I., 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One 3, e3042. doi:10.1371/journal.pone.0003042

Gomez-Alvarez, V., Revetta, R.P., Domingo, J.W.S., 2012. Metagenomic analyses of drinking water receiving different disinfection treatments. Appl. Environ. Microbiol. 78, 6095–6102. doi:10.1128/AEM.01018-12

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317. doi:10.1038/ismej.2009.72

Gómez-Pastor, R., Pérez-Torrado, R., Cabiscol, E., Ros, J., Matallana, E., 2010. Reduction of oxidative cellular damage by overexpression of the thioredoxin *TRX2* gene improves yield and quality of wine yeast dry active biomass. Microb. Cell Fact. 9, 9. doi:10.1186/1475-2859-9-9

Gray, M.J., Wholey, W.-Y., Jakob, U., 2013. Bacterial responses to reactive chlorine species. Annu. Rev. Microbiol. 67, 141–60. doi:10.1146/annurev-micro-102912-142520

Guo, J., Peng, Y., Ni, B.-J., Han, X., Fan, L., Yuan, Z., 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. Microb. Cell Fact. 14, 33. doi:10.1186/s12934-015-0218-4

Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L., Rolain, J.-M., 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob. Agents Chemother. 58, 212–220. doi:10.1128/AAC.01310-13

Hall, C.C., Watkins, J.D., Georgopapadakou, N.H., 1989. Effects of elfamycins on elongation factor Tu from *Escherichia coli* and *Staphylococcus aureus*. Antimicrob. Agents Chemother. 33, 322–5.

Hamidian, M., Holt, K.E., Pickard, D., Hall, R.M., 2016. A small *Acinetobacter* plasmid carrying the *tet39* tetracycline resistance determinant. J. Antimicrob. Chemother. 71, 269–71. doi:10.1093/jac/dkv293

Hong, P.-Y., Hwang, C., Ling, F., Andersen, G.L., LeChevallier, M.W., Liu, W.-T., 2010. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. Appl. Environ. Microbiol. 76, 5631–5635. doi:10.1128/AEM.00281-10

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M., Brown, C.T., 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proc. Natl. Acad. Sci. 111, 4904–4909. doi:10.1073/pnas.1402564111

Huang, J.-J., Hu, H.-Y., Tang, F., Li, Y., Lu, S.-Q., Lu, Y., 2011. Inactivation and reactivation of antibiotic-resistant bacteria by chlorination in secondary effluents of a municipal wastewater treatment plant. Water Res. 45, 2775–2781. doi:10.1016/j.watres.2011.02.026

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. doi:10.1101/gr.5969107

Ilmberger, N., Güllert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., Alawi, M., Poehlein, A., Chow, J., Turaev, D., Rattei, T., Schmeisser, C., Salomon, J., Olsen, P.B., Daniel, R., Grundhoff, A., Borchert, M.S., Streit, W.R., 2014. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One 9, e106707. doi:10.1371/journal.pone.0106707

Jeffries, T.C., Seymour, J.R., Gilbert, J.A., Dinsdale, E.A., Newton, K., Leterme, S.S.C., Roudnew, B., Smith, R.J., Seuront, L., Mitchell, J.G., 2011. Substrate type determines metagenomic profiles from diverse chemical habitats. PLoS One 6, e25173. doi:10.1371/journal.pone.0025173

Ju, F., Guo, F., Ye, L., Xia, Y., Zhang, T., 2014. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. Environ. Microbiol. Rep. 6, 80–89. doi:10.1111/1758-2229.12110

Karumathil, D.P., Yin, H.-B., Kollanoor-Johny, A., Venkitanarayanan, K., 2014. Effect of chlorine exposure on the survival and antibiotic gene expression of multidrug resistant *Acinetobacter baumannii* in water. Int. J. Environ. Res. Public Health 11, 1844–1854. doi:10.3390/ijerph110201844

Kristiansson, E., Fick, J., Janzon, A., Grabic, R., Rutgersson, C., Weijdegård, B., Söderström, H., Larsson, D.G.J., 2011. Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One 6, e17038. doi:10.1371/journal.pone.0017038

Kuhns, L.G., Wang, G., Maier, R.J., 2015. Comparative roles of the two *Helicobacter pylori* Thioredoxins in preventing macromolecule damage. Infect. Immun. 83, 2935–43. doi:10.1128/IAI.00232-15

Kümmerer, K., 2009. Antibiotics in the aquatic environment – a review – Part II. Chemosphere 75, 435–441. doi:10.1016/j.chemosphere.2008.12.006

Lim, Y.W., Evangelista, J.S., Schmieder, R., Bailey, B., Haynes, M., Furlan, M., Maughan, H., Edwards, R., Rohwer, F., Conrad, D., 2014. Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. J. Clin. Microbiol. 52, 425–437. doi:10.1128/JCM.02204-13

Liu, B., Pop, M., 2009. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res. 37, D443–437. doi:10.1093/nar/gkn656

Lupo, A., Coyne, S., Berendonk, T.U., 2012. Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. Front. Microbiol. 3. doi:10.3389/fmicb.2012.00018

Maiese, W.M., Lechevalier, M.P., Lechevalier, H.A., Korshalla, J., Goodman, J., Wildey, M.J., Kuck, N., Conner, S.D., Greenstein. M., 1989. LL-E19020α and β, animal growth promoting antibiotics: Taxonomy, Fermentation and Biological activity. J. Antibiot. 42, 1489–1493.

Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., Hou, J., Mu, Q., Alvarez, P.J.J., 2015. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Res. 85, 458–466. doi:10.1016/j.watres.2015.09.010

Mason, O.U., Scott, N.M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., Bouskill, N.J., Prestat, E., Borglin, S., Joyner, D.C., Fortney, J.L., Jurelevicius, D., Stringfellow, W.T., Alvarez-Cohen, L., Hazen, T.C., Knight, R., Gilbert, J.A., Jansson, J.K., 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J. 8, 1464–1475. doi:10.1038/ismej.2013.254

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., Pascale, G. De, Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J. V, Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., Wright, G.D., 2013. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57, 3348–3357. doi:10.1128/AAC.00419-13

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386. doi:10.1186/1471-2105-9-386

Miele, A., Goldstein, B.P., Bandera, M., Jarvis, C., Resconi, A., Williams, R.J., 1994. Differential susceptibilities of enterococcal species to elfamycin antibiotics. J. Clin. Microbiol. 32, 2016–8.

Miura, Y., Hiraiwa, M.N., Ito, T., Itonaga, T., Watanabe, Y., Okabe, S., 2007. Bacterial community structures in MBRs treating municipal wastewater: relationship between community stability and reactor performance. Water Res. 41, 627–37. doi:10.1016/j.watres.2006.11.005

Müller, S., Becker, K., Bergmann, B., Schirmer, R.H., Walter, R.D., 1995. Plasmodium falciparum glutathione reductase exhibits sequence similarities with the human host enzyme in the core structure but differs at the ligand-binding sites. Mol. Biochem. Parasitol. 74, 11–8.

Murray, G.E., Tobin, R.S., Junkins, B., Kushner, D.J., 1984. Effect of chlorination on antibiotic resistance profiles of sewage-related bacteria. Appl. Environ. Microbiol. 48, 73–77.

Odjadjare, E.C., Olaniran, A.O., 2015. Prevalence of antimicrobial resistant and virulent *Salmonella* spp. in treated effluent and receiving aquatic milieu of wastewater treatment plants in Durban, South Africa. Int. J. Environ. Res. Public Health 12, 9692–9713. doi:10.3390/ijerph120809692

Olaniran, A.O., Naidoo, S., Pillay, B., 2012. Surveillance of invasive bacterial pathogens and human enteric viruses in wastewater final effluents and receiving water bodies – a case study from Durban, South Africa. CLEAN – Soil, Air, Water 40, 681–691. doi:10.1002/clen.201100023

Olaniran, A.O., Nzimande, S.B.T., Mkize, N.G., 2015. Antimicrobial resistance and virulence signatures of *Listeria* and *Aeromonas* species recovered from treated wastewater effluent and receiving surface water in Durban, South Africa. BMC Microbiol. 15, 234. doi:10.1186/s12866-015-0570-x

Parks, D.H., Beiko, R.G., 2010. Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26, 715–21. doi:10.1093/bioinformatics/btq041

Poitelon, J.-B., Joyeux, M., Welté, B., Duguet, J.-P., Prestel, E., Lespinet, O., DuBow, M.S., 2009. Assessment of phylogenetic diversity of bacterial microflora in drinking water using serial analysis of ribosomal sequence tags. Water Res. 43, 4197–206. doi:10.1016/j.watres.2009.07.020

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K. Le, Leclerc, M., Maguin, E., Minardi, R.M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65. doi:10.1038/nature08821

R Development Core Team, 2007. R: A language and environment for statistical computing. Vienna: R Development Core Team.

Ram, S., Vajpayee, P., Shanker, R., 2008. Contamination of potable water distribution systems by multiantimicrobial-resistant enterohemorrhagic *Escherichia coli*. Environ. Health Perspect. 116, 448–452. doi:10.1289/ehp.10809

Revetta, R.P., Matlib, R.S., Santo Domingo, J.W., 2011. 16S rRNA gene sequence analysis of drinking water using RNA and DNA extracts as targets for clone library development. Curr. Microbiol. 63, 50–59. doi:10.1007/s00284-011-9938-9

Rho, M., Tang, H., Ye, Y., 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. 38, e191. doi:10.1093/nar/gkq747

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I., Fatta-Kassinos, D., 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci. Total Environ. 447, 345–360. doi:10.1016/j.scitotenv.2013.01.032

Russel, M., Model, P., 1988. Sequence of thioredoxin reductase from *Escherichia coli*. Relationship to other flavoprotein disulfide oxidoreductases. J. Biol. Chem. 263, 9015–9.

Ryu, J.-H., Beuchat, L.R., 2005. Biofilm formation by *Escherichia coli* O157:H7 on stainless steel: effect of exopolysaccharide and curli production on its resistance to chlorine. Appl. Environ. Microbiol. 71, 247–54. doi:10.1128/AEM.71.1.247-254.2005

Saby, S., Leroy, P., Block, J.C., 1999. *Escherichia coli* resistance to chlorine and glutathione synthesis in response to oxygenation and starvation. Appl. Environ. Microbiol. 65, 5600–3.

Scaria, J., Chandramouli, U., Verma, S.K., 2005. Antibiotic Resistance Genes Online (ARGO): a database on vancomycin and beta-lactam resistance genes. Bioinformation 1, 5–7.

Schmieder, R., Edwards, R., 2012. Insights into antibiotic resistance through metagenomic approaches. Future Microbiol. 7, 73–89. doi:10.2217/fmb.11.135

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl. Environ. Microbiol. 79, 2906–13. doi:10.1128/AEM.03448-12

Shen S. Y., Fulthorpe R., 2015. Seasonal variation of bacterial endophytes in urban trees. Front. Microbiol. 6, 427. doi:10.3389/fmicb.2015.00427

Shi, P., Jia, S., Zhang, X.-X., Zhang, T., Cheng, S., Li, A., 2013. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res. 47, 111–120. doi:10.1016/j.watres.2012.09.046

Shimizu, K., 2013. Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. Metabolites 4, 1–35. doi:10.3390/metabo4010001

Silva, C.C., Jesus, E.C., Torres, A.P.R., Sousa, M.P., Santiago, V.M.J., Oliveira, V.M., 2010a. Investigation of bacterial diversity in membrane bioreactor and conventional activated sludge processes from petroleum refineries using phylogenetic and statistical approaches. J. Microbiol. Biotechnol. 20, 447–59.

Silva, C.C., Viero, A.F., Dias, A.C.F., Andreote, F.D., Jesus, E.C., De Paula, S.O., Torres, A.P.R., Santiago, V.M.J., Oliveira, V.M., 2010b. Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load. J. Microbiol. Biotechnol. 20, 21–9.

Smirnova, G. V, Krasnykh, T.A., Oktyabrsky, O.N., 2001. Role of glutathione in the response of *Escherichia coli* to osmotic stress. Biochem. Biokhimii□a□ 66, 973–8.

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., Newton, K., Brown, M.H., Mitchell, J.G., 2012. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. Environ. Microbiol. 14, 240–53. doi:10.1111/j.1462-2920.2011.02614.x

Su, J.-Q., Wei, B., Ou-Yang, W.-Y., Huang, F.-Y., Zhao, Y., Xu, H.-J., Zhu, Y.-G., 2015. Antibiotic resistome and its association with bacterial communities during sewage sludge composting. Environ. Sci. Technol. 49, 7356–7363. doi:10.1021/acs.est.5b01012

Sund, C.J., Rocha, E.R., Tzianabos, A.O., Tzinabos, A.O., Wells, W.G., Gee, J.M., Reott, M.A., O'Rourke, D.P., Smith, C.J., 2008. The *Bacteroides fragilis* transcriptome response to oxygen and H_2O_2: the role of OxyR and its effect on survival and virulence. Mol. Microbiol. 67, 129–42. doi:10.1111/j.1365-2958.2007.06031.x

Vogeley, L., Palm, G.J., Mesters, J.R., Hilgenfeld, R., 2001. Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. J. Biol. Chem. 276, 17149–55. doi:10.1074/jbc.M100017200

Wang, S., Deng, K., Zaremba, S., Deng, X., Lin, C., Wang, Q., Tortorello, M. Lou, Zhang, W., 2009. Transcriptomic response of *Escherichia coli* O157:H7 to oxidative stress. Appl. Environ. Microbiol. 75, 6110–23. doi:10.1128/AEM.00914-09

Wang, Z., Zhang, X.-X., Huang, K., Miao, Y., Shi, P., Liu, B., Long, C., Li, A., 2013. Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. PLoS One 8, e76079. doi:10.1371/journal.pone.0076079

Wellington, E.M.H., Boxall, A.B.A., Cross, P., Feil, E.J., Gaze, W.H., Hawkey, P.M., Johnson-Rollings, A.S., Jones, D.L., Lee, N.M., Otten, W., Thomas, C.M., Williams, A.P., 2013. The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. Lancet Infect. Dis. 13, 155–165. doi:10.1016/S1473-3099(12)70317-1

WHO, 2014. Antimicrobial resistance: global report on surveillance 2014.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P., Paczian, T., Trimble, W.L., Bagchi, S., Grama, A., Chaterji, S., Meyer, F., 2015. The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Res. 44, D590–4. doi:10.1093/nar/gkv1322

Wolf, H., Chinali, G., Parmeggiani, A., 1974. Kirromycin, an inhibitor of protein biosynthesis that acts on elongation factor Tu. Proc. Natl. Acad. Sci. U. S. A. 71, 4910–4.

Xi, C., Zhang, Y., Marrs, C.F., Ye, W., Simon, C., Foxman, B., Nriagu, J., 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Appl. Environ. Microbiol. 75, 5714–5718. doi:10.1128/AEM.00382-09

Yadav, T.C., Khardenavis, A.A., Kapley, A., 2014. Shifts in microbial community in response to dissolved oxygen levels in activated sludge. Bioresour. Technol. 165, 257–64. doi:10.1016/j.biortech.2014.03.007

Yang, J., 2014. Viral pathogens in clinical samples by use of a metagenomic approach, in: Nelson, K.E. (Ed.), Encyclopedia of Metagenomics. Springer New York, pp. 1–6.

Ye, L., Zhang, T., 2013. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Appl. Microbiol. Biotechnol. 97, 2681–2690. doi:10.1007/s00253-012-4082-4

Yu, K., Zhang, T., 2012. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7, e38183. doi:10.1371/journal.pone.0038183

Yu, K., Zhang, T., 2013. Construction of customized sub-databases from NCBI-nr database for rapid annotation of huge metagenomic datasets using a combined BLAST and MEGAN approach. PLoS One 8, e59831. doi:10.1371/journal.pone.0059831

Zhang, T., Zhang, X.-X., Ye, L., 2011. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One 6, e26041. doi:10.1371/journal.pone.0026041

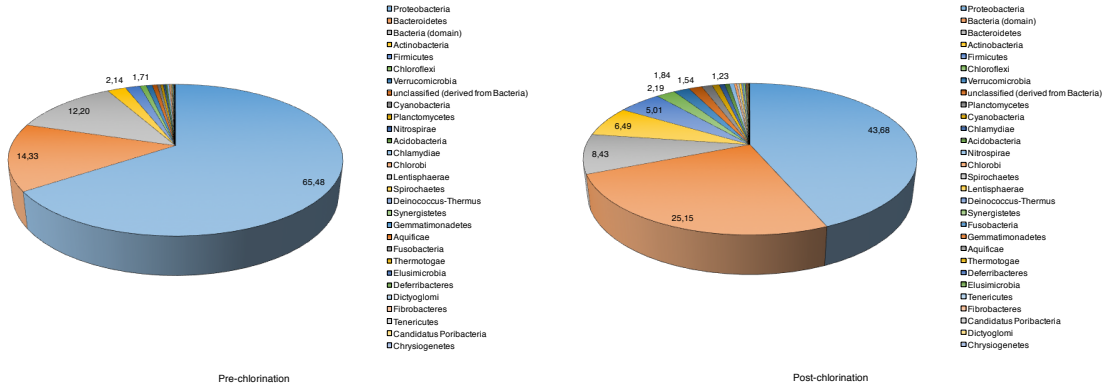
Zhang, Y., Zhuang, Y., Geng, J., Ren, H., Zhang, Y., Ding, L., Xu, K., 2015. Inactivation of antibiotic resistance genes in municipal wastewater effluent by chlorination and sequential UV/chlorination disinfection. Sci. Total Environ. 512-513, 125–132. doi:10.1016/j.scitotenv.2015.01.028

Zheng, M., Doan, B., Schneider, T.D., Storz, G., 1999. OxyR and SoxRS regulation of *fur*. J. Bacteriol. 181, 4639–43.

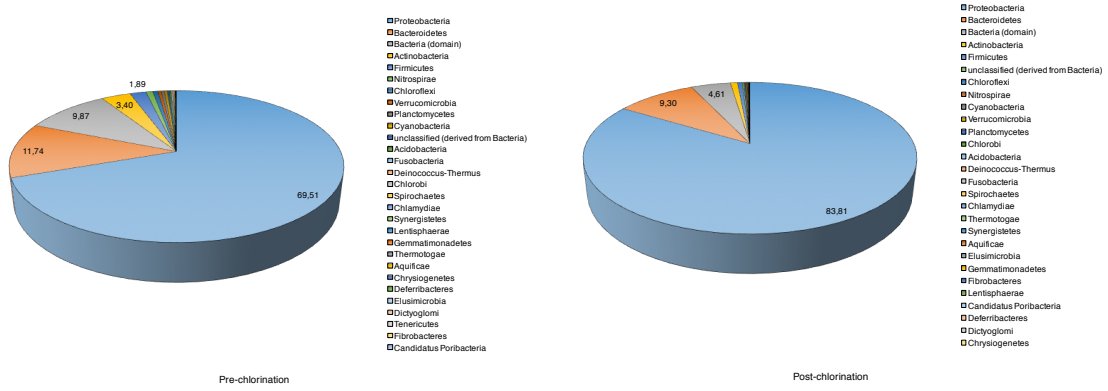
Zheng, W., Zhang, Z., Liu, C., Qiao, Y., Zhou, D., Qu, J., An, H., Xiong, M., Zhu, Z., Zhao, X., 2015. Metagenomic sequencing reveals altered metabolic pathways in the oral microbiota of sailors during a long sea voyage. Sci. Rep. 5, 9131. doi:10.1038/srep09131

Zhou, C.E., Smith, J., Lam, M., Zemla, A., Dyer, M.D., Slezak, T., 2007. MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res. 35, D391–4. doi:10.1093/nar/gkl791

3.8 Figures



(a)



(b)

Figure 3.1: Relative abundances of dominant phyla of the pre- and post-chlorinated samples for (**a**) Plant A and (**b**) Plant B annotated by MG-RAST. Relative abundance represents the number of reads affiliated with that taxon divided by the total reads assigned for the bacterial domain, and the phylum (or domain) making up more than 1% of the population in the sample is indicated on the chart as percentages.

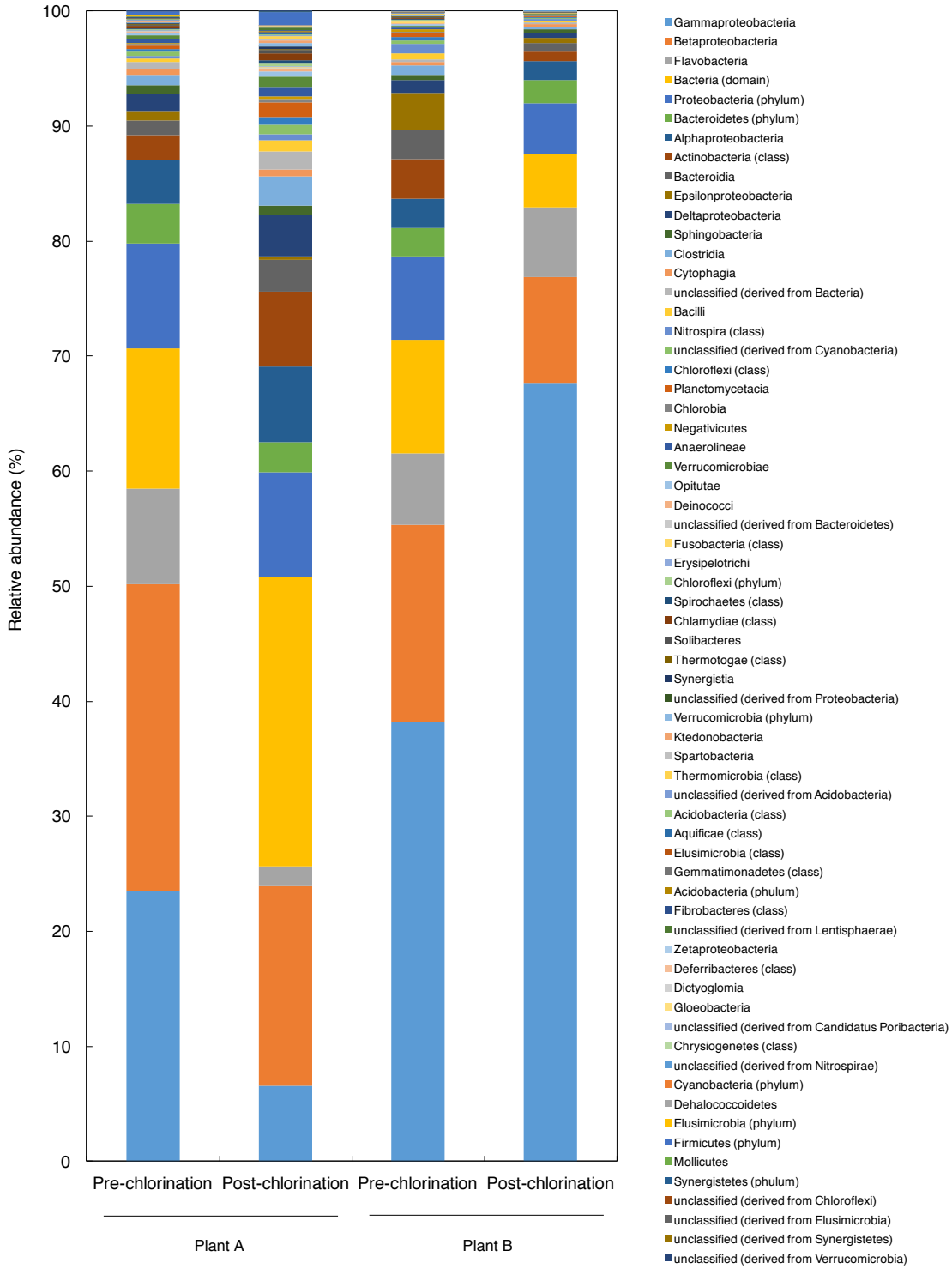


Figure 3.2: Relative abundances of dominant class in the pre- and post-chlorinated wastewater samples from Plant A and Plant B annotated by MG-RAST. Relative abundance represents the number of reads affiliated with that taxon divided by the total reads assigned for the bacterial domain.

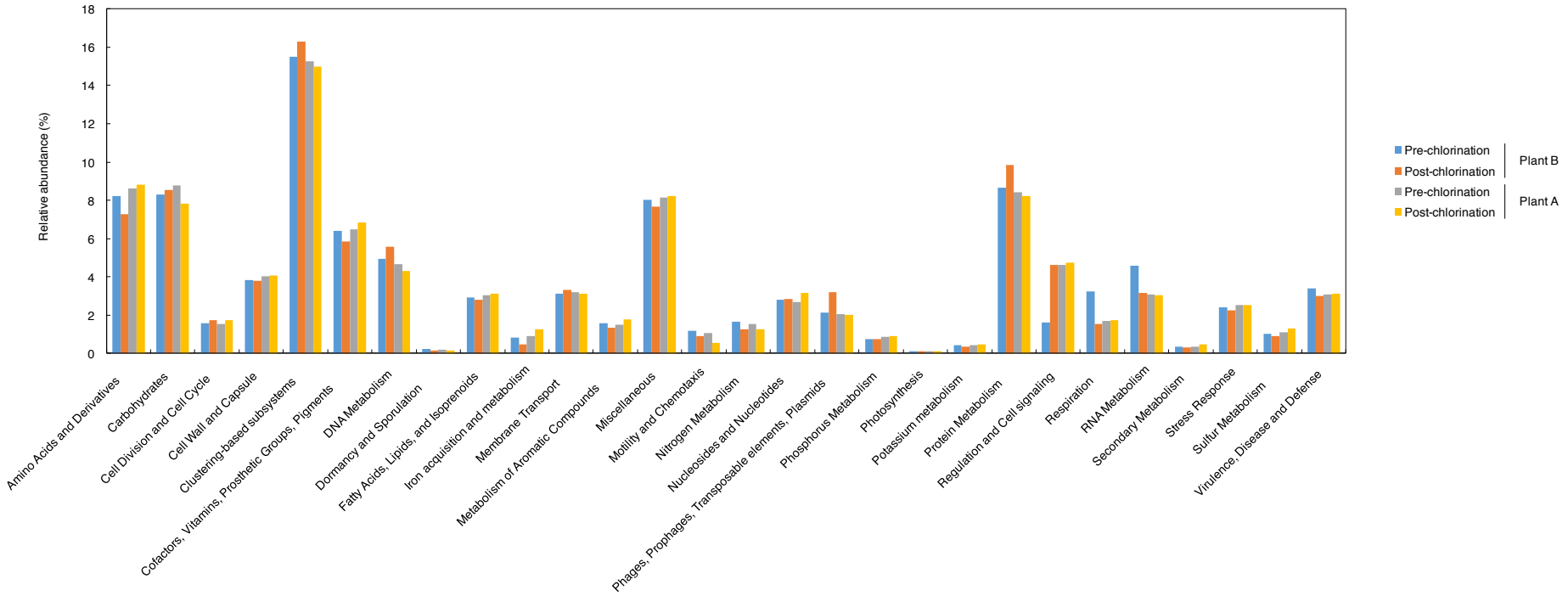


Figure 3.3: Comparison on the relative abundance of functional annotation indicating the variation of bacterial metabolic structure at level 1 SEED subsystems. Relative abundance represents the number of reads affiliated with that taxon divided by the total reads assigned for the bacterial domain.

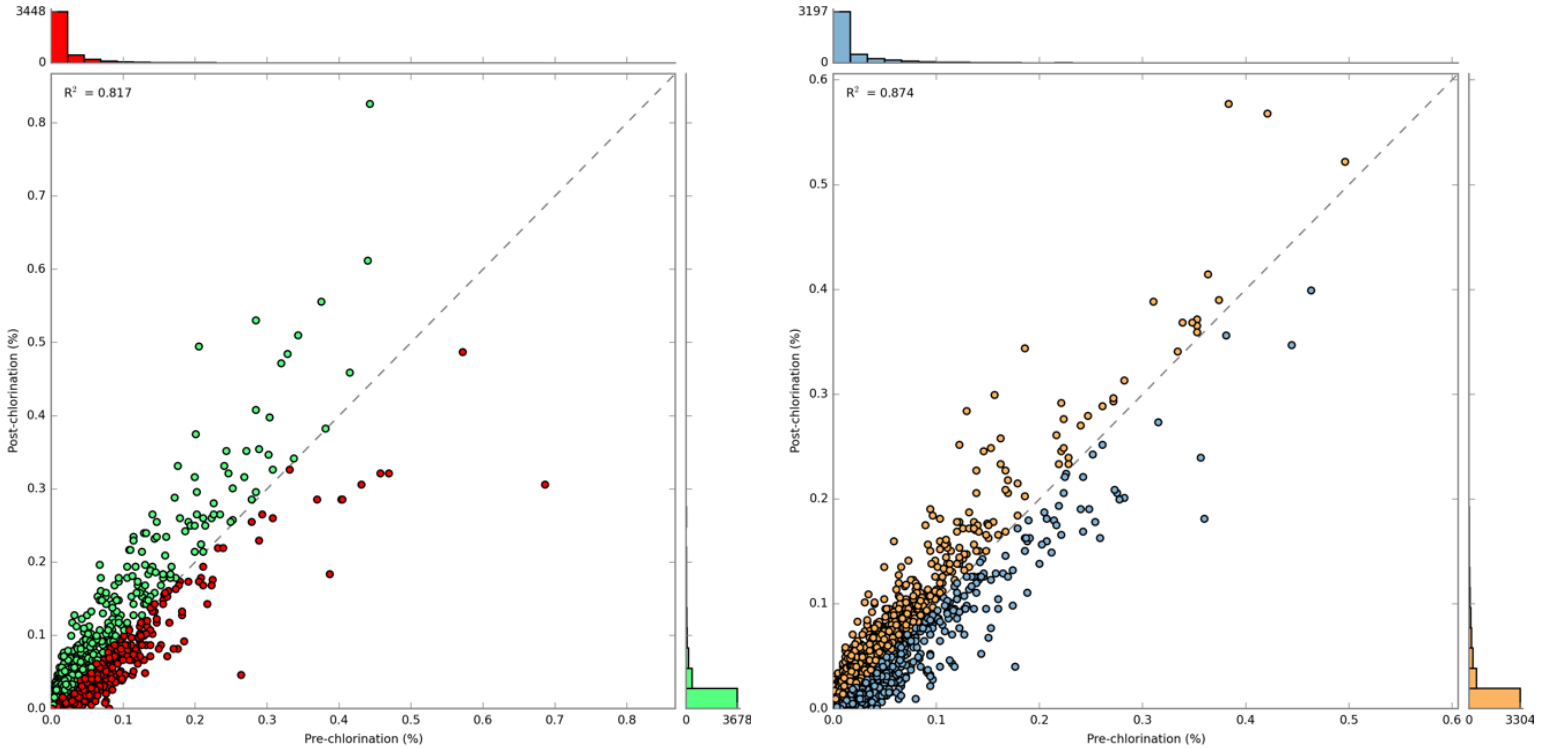
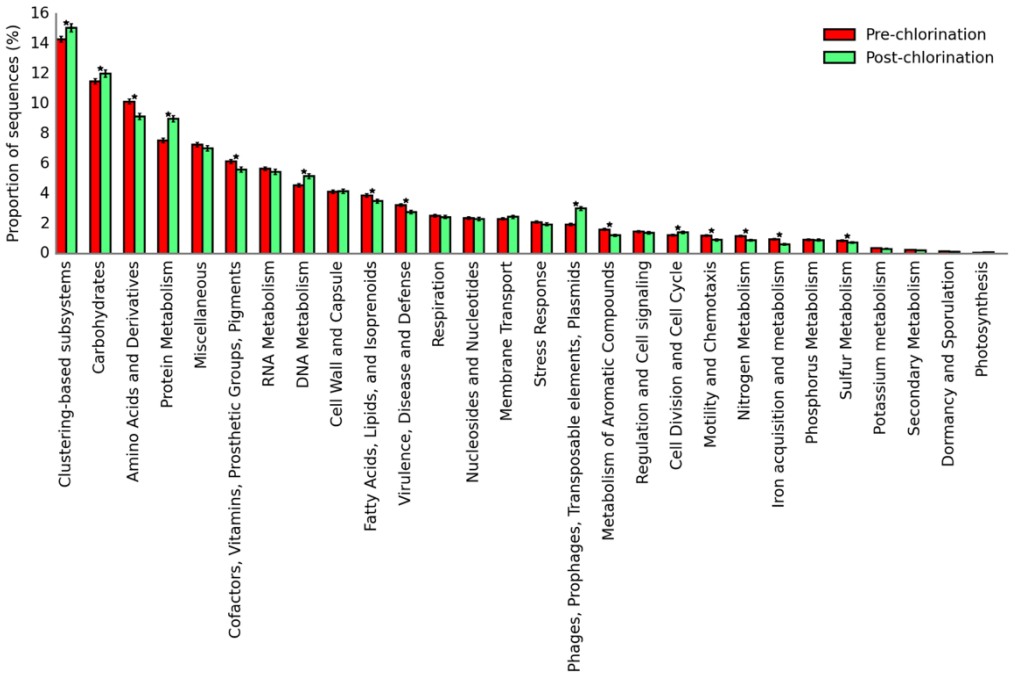
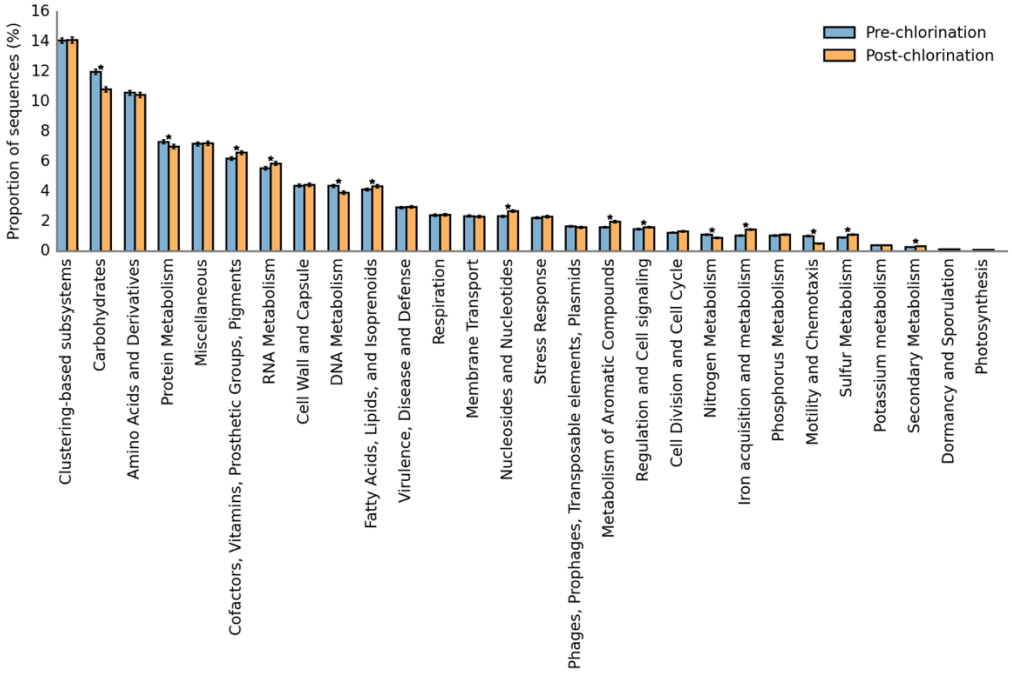


Figure 3.4: Comparative level 1 SEED subsystem profile of microbial communities in the pre- and post-chlorinated wastewater. Statistical analysis using STAMP showed overall negative correlations in (**a**) Plant A, and positive correlations in (**b**) Plant B. Significant differences (q<0.05) in abundance between each subsystems before and after disinfection treatment in (**c**) Plant A and (**d**) Plant B annotated by MG-RAST are indicated in asterisks.



(**c**)



(**d**)

Figure 3.4 (cont.): Comparative level 1 SEED subsystem profile of microbial communities in the pre- and post-chlorinated wastewater. Statistical analysis using STAMP showed overall negative correlations in (**a**) Plant A, and positive correlations in (**b**) Plant B. Significant differences (q<0.05) in abundance between each subsystems before and after disinfection treatment in (**c**) Plant A and (**d**) Plant B annotated by MG-RAST are indicated in asterisks.

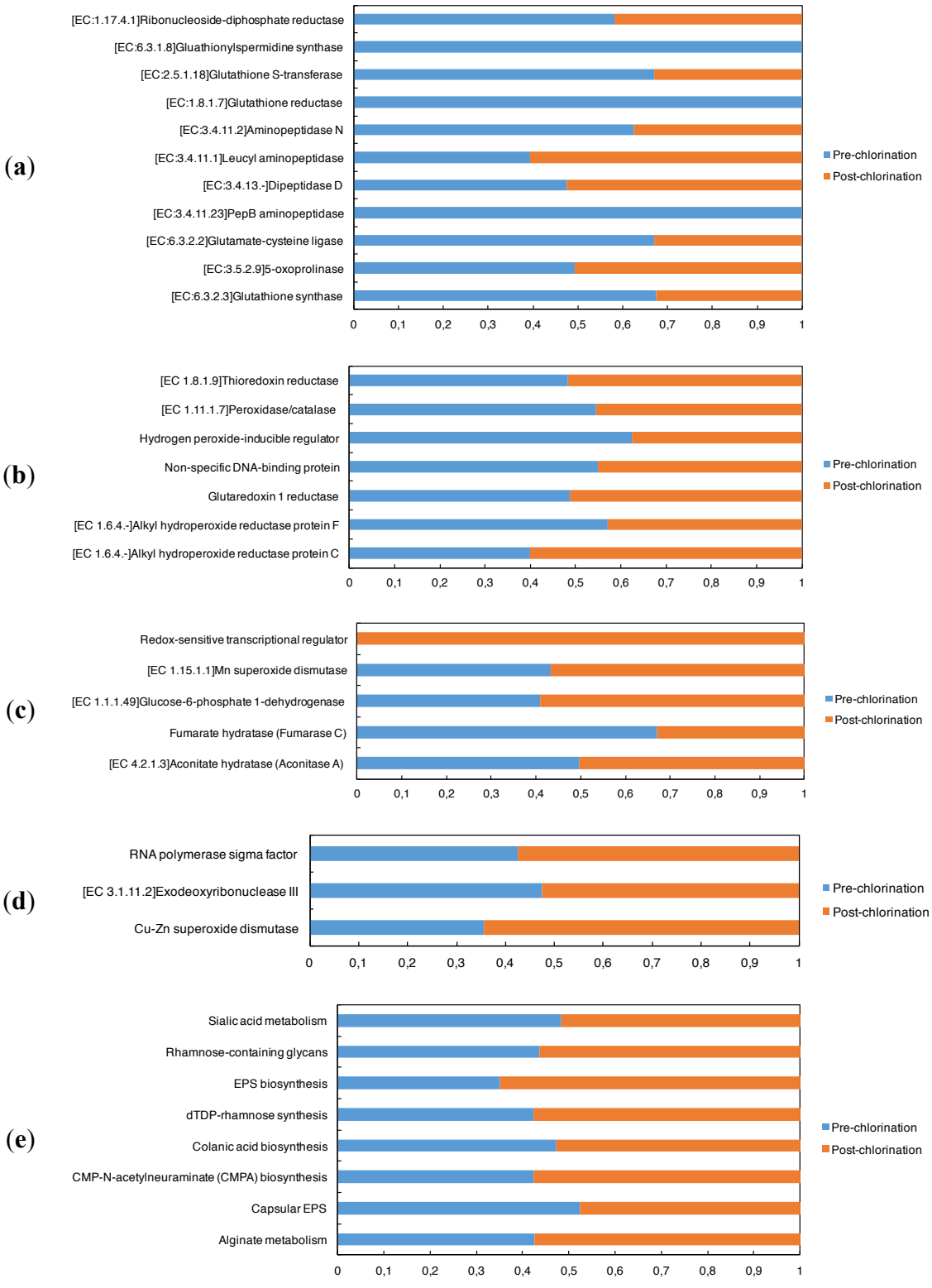


Figure 3.5: Normalized relative abundance of annotated reads conferring resistance to chlorine disinfection under (**a**) Glutathion pathway assigned in KO (**b**) OxyR system, (**c**) SoxRS system, (**d**) RpoS regulated genes and (**e**) exopolysaccharide (EPS) synthesis pathways assigned in SS of pre- and post-chlorination disinfection wastewaters of Plant A. Relative abundance was calculated from the number of reads annotated in either KO or SS.

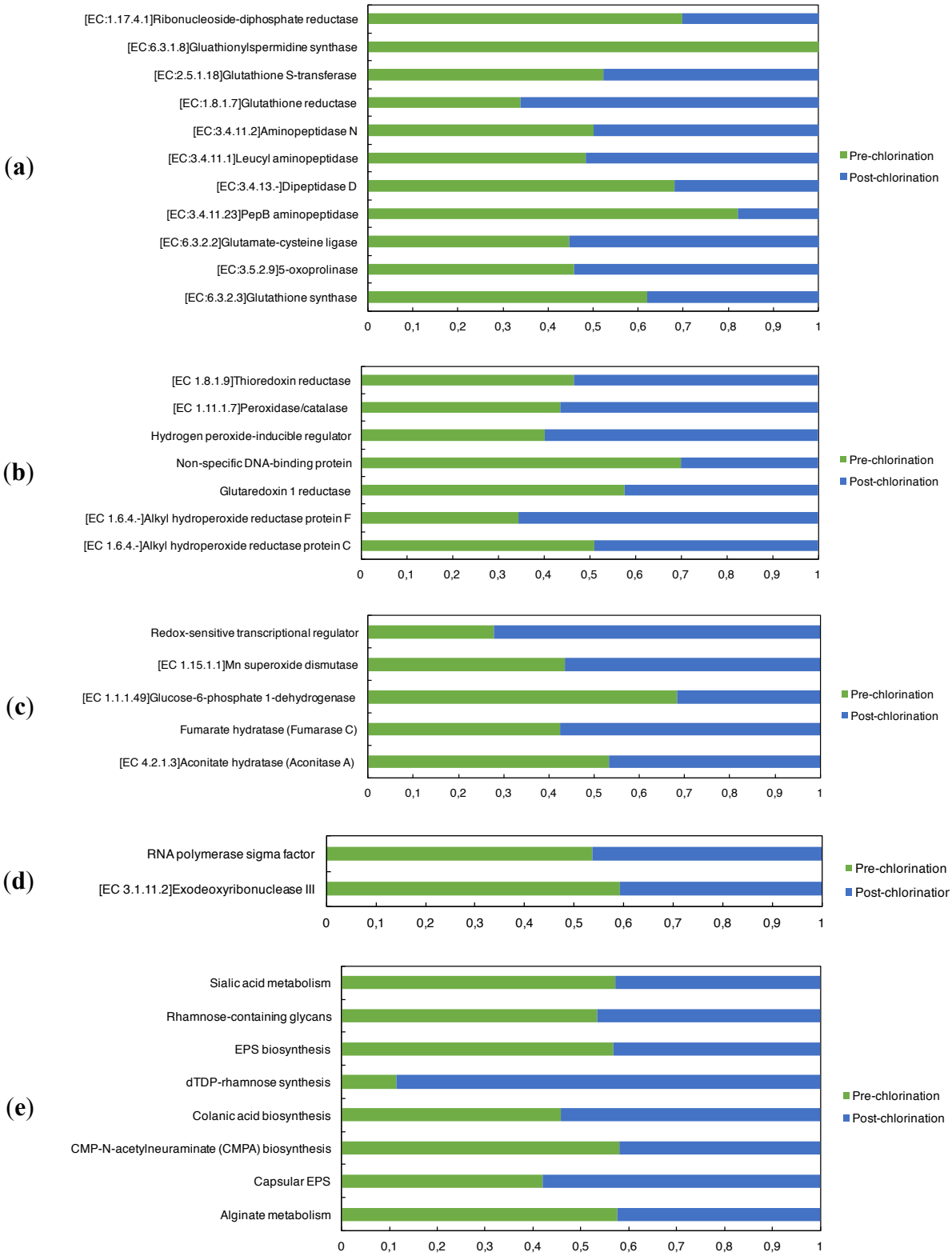


Figure 3.6: Normalized relative abundance of annotated reads conferring resistance to chlorine disinfection under (**a**) Glutathion pathway assigned in KO (**b**) OxyR system, (**c**) SoxRS system, (**d**) RpoS regulated genes and (**e**) exopolysaccharide (EPS) synthesis pathways assigned in SS of pre- and post-chlorination disinfection wastewaters of Plant B. Relative abundance was calculated from the number of reads annotated in either KO or SS.

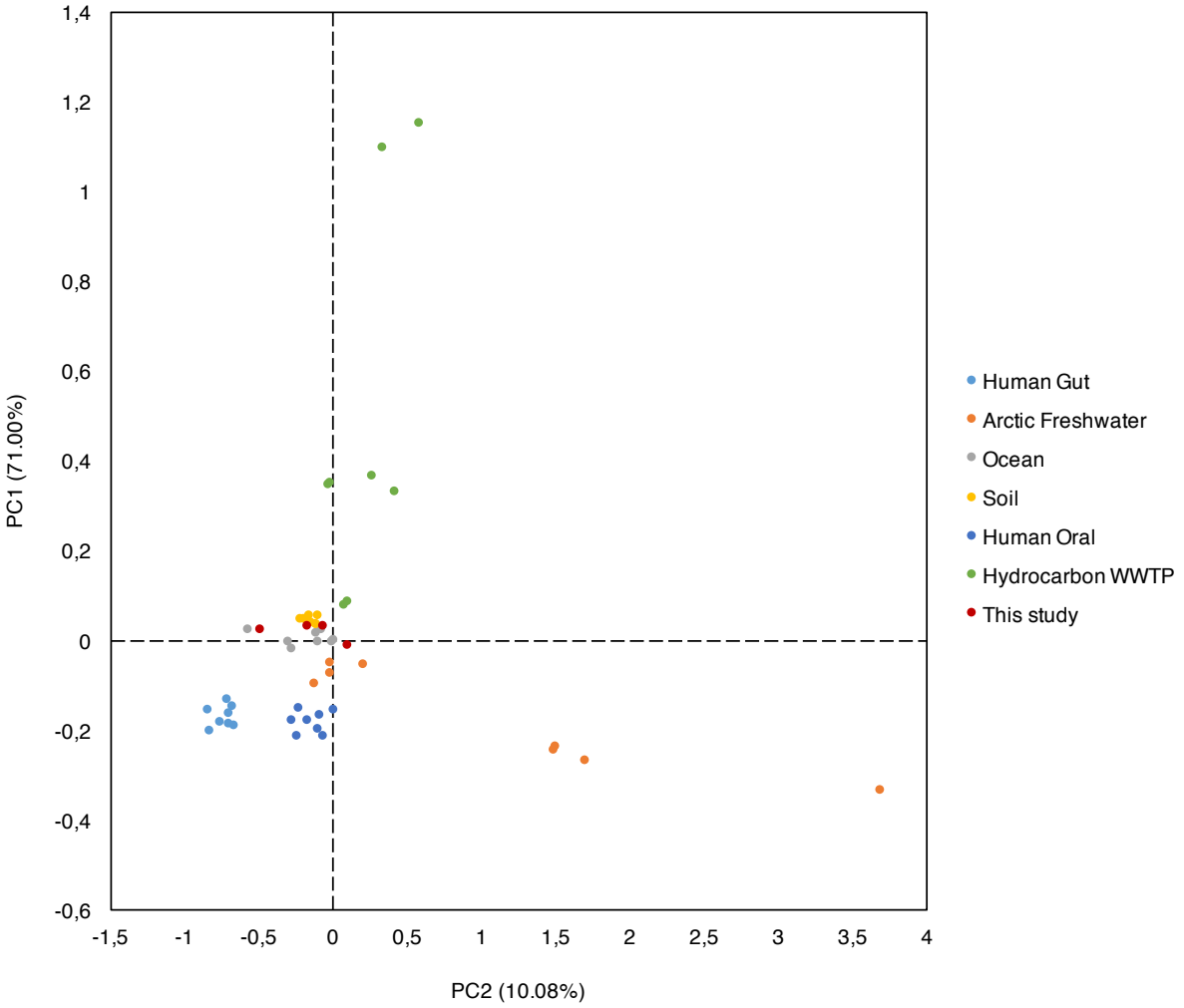


Figure 3.7: Principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity coefficients of 6 different ecosystems (human oral, human gut, ocean, soil freshwater, wastewater treatment plants) with normalized amount of annotated reads in the SEED subsystems. Metagenomes of the different ecosystems were analysed in MG-RAST to avoid variations in analysis using the publicly available data on MG-RAST (their information is provided in Table S3.2).

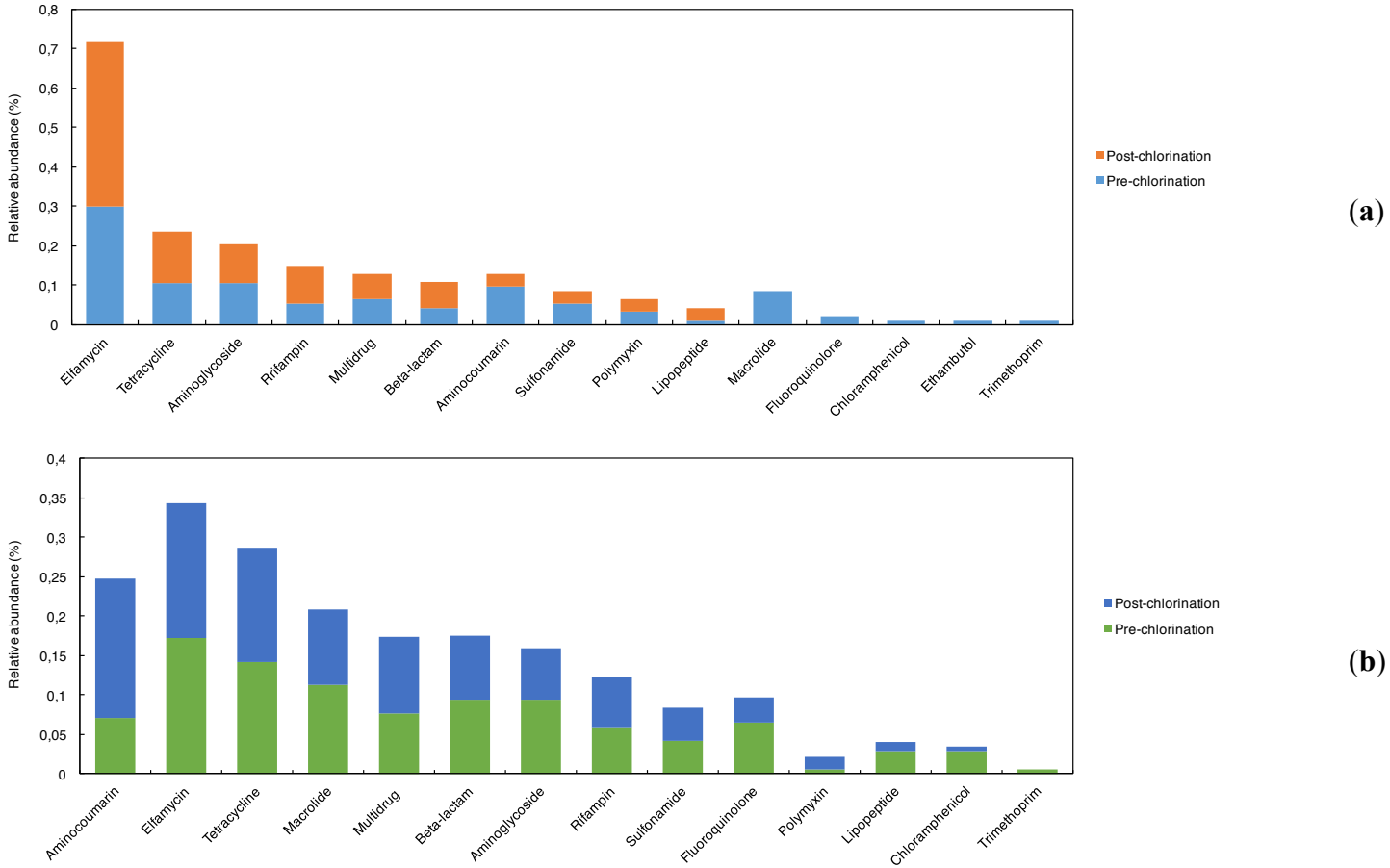
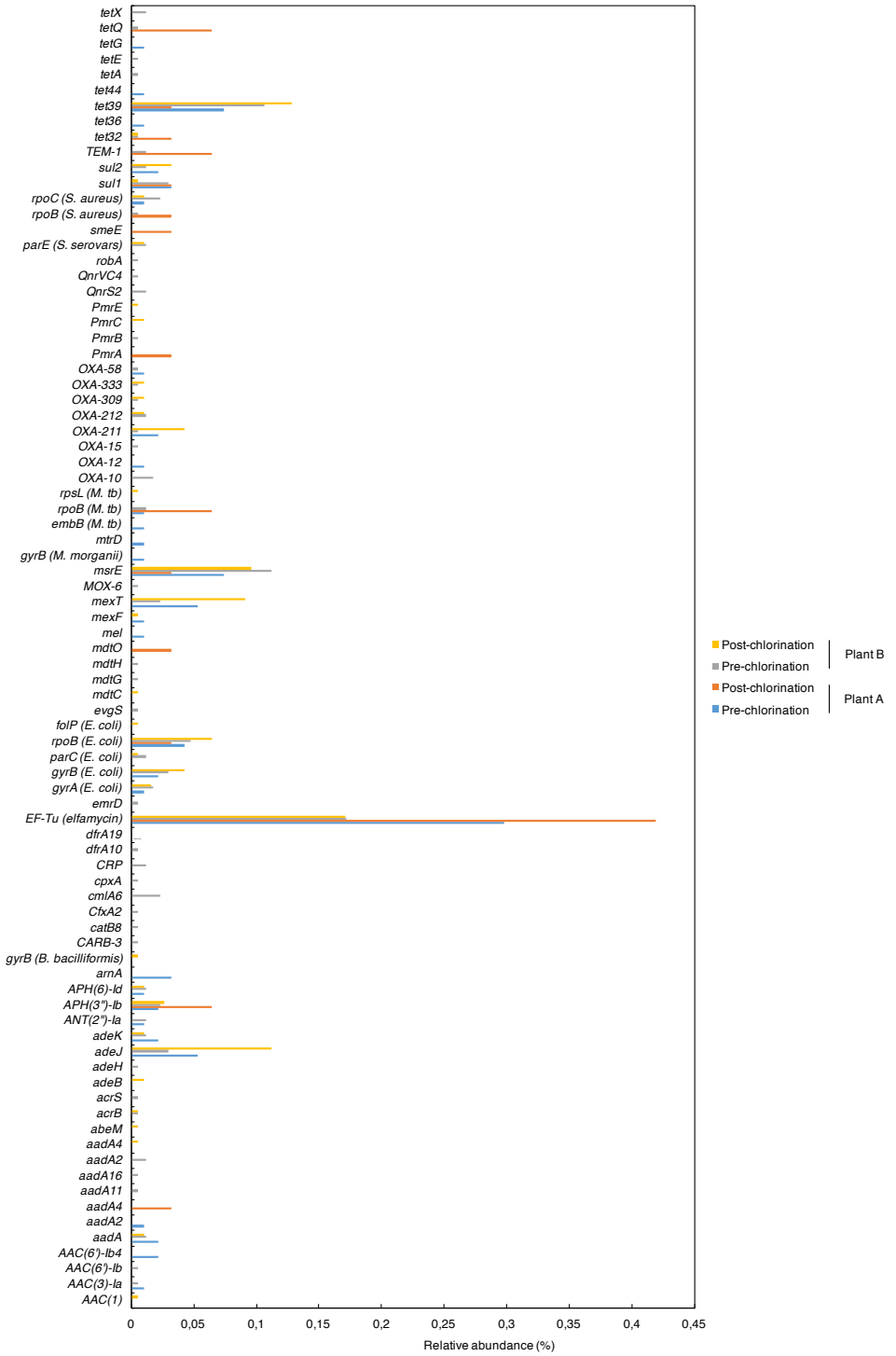


Figure 3.8: Comparison of the relative abundance of antibiotic resistance gene (ARGs) classes observed from bacterial communities (pre- and post-chlorination) in (**a**) Plant A and (**b**) Plant B and the diversity of specific ARGs in (**c**) Plant A and Plant B. Relative abundance values were calculated by dividing the number of assigned ARGs to the total number of ARGs detected. ARGs were annotated by BLASTx against the CARD protein database. (*Staphylococcus aureus*, *S. aureus*; *Salmonella enterica*, *S. serovars*; *Mycobacterium tuberculosis*, *M. tb*; *Escherichia coli*, *E. coli; Bartonella bacilliformis, B. bacilliformis*).



(**c**)

Figure 3.8 (cont.): Comparison of the relative abundance of antibiotic resistance gene (ARGs) classes observed from bacterial communities (pre- and post-chlorination) in (**a**) Plant A and (**b**) Plant B and the diversity of specific ARGs in (**c**) Plant A and Plant B. Relative abundance values were calculated by dividing the number of assigned ARGs to the total number of ARGs detected. ARGs were annotated by BLASTx against the CARD protein database. (*Staphylococcus aureus*, *S. aureus*; *Salmonella enterica*, *S. serovars*; *Mycobacterium tuberculosis*, *M. tb*; *Escherichia coli*, *E. coli; Bartonella bacilliformis, B. bacilliformis*).

3.9 Supplementary material

Table S3.1: Characterization of the 454 pyrosequenced libraries from the pre- and post-chlorinated wastewater samples of two full-scale urban wastewater treatment plants in Durban, South Africa.

Characteristics	Plant A		Plant B	
	Pre-chlorination	**Post-chlorination**	**Pre-chlorination**	**Post-chlorination**
# raw reads	237,014	227,021	270,395	180,557
Total size (bp)	151,058,869	143,321,109	73,708,040	110,996,053
Average reads (bp)	637	631	272	614
# reads uploaded to MG-RAST	237,014	227,021	270,395	180,557
MG-RAST QC				
# reads before QC	237,014	227,021	270,395	180,557
Total size before QC (bp)	151,058,869	143,321,109	73,708,040	110,996,053
Average length before QC (bp)	637	631	272	614
# removed during dereplication (bp)	11,741	11,511	13,020	2,031
# after QC	203,253	194,388	217,147	147,749
Total size after QC (bp)	42,025,286	41,085,936	44,369,381	27,636,500
Average length after QC (bp)	207	211	204	187
# predicted ORFs	184,950	175,950	198,687	131,523
# identified RNA	32,915	32,299	35,691	23,065
Functional Annotation				
# of SEED entries	6797	6041	7034	5759
# of SEED subsystem level 2	167	165	166	162
# of SEED subsystem level 3	978	946	1,005	928
# of KEGG entries	1961	1881	2040	1607
# of COG entries	3012	2769	3188	2622

Table S3.2: Basic information summary of 48 publicly available metagenomic datasets from MG-RAST in six ecosystems (human gut, human oral, ocean, soil, freshwater and wastewater treatment plants) used in this study.

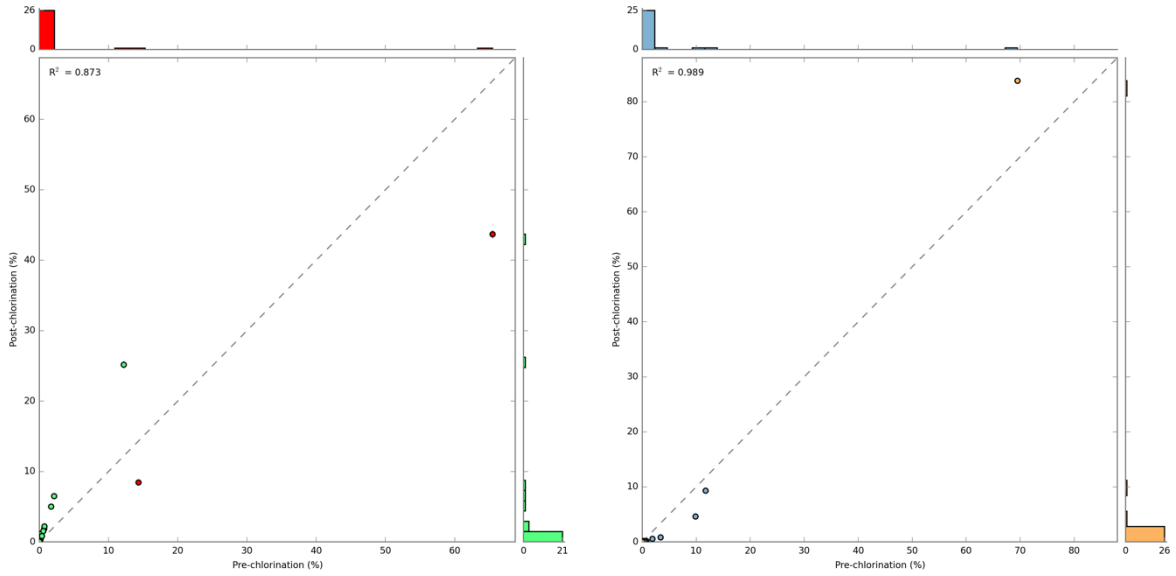
Label/Biomes	MG-RAST ID	Metagenome size (bp)	# Reads	Habitat Type	Location	Sequencing Method
Human Oral_1	4447943.3	142,374,233	339,503	human-associated	Valencia, Spain	454
Human Oral_2	4447192.3	77,538,485	204,218	human-associated	Valencia, Spain	454
Human Oral_3	4447103.3	203,711,161	464,594	human-associated	Valencia, Spain	454
Human Oral_4	4447102.3	100,125,112	244,881	human-associated	Valencia, Spain	454
Human Oral_5	4447101.3	129,851,692	295,072	human-associated	Valencia, Spain	454
Human Oral_6	4447971.3	37,519,874	97,722	human-associated	Valencia, Spain	454
Human Oral_7	4447970.3	27,669,924	70,503	human-associated	Valencia, Spain	454
Human Oral_8	4447903.3	123,266,763	306,74	human-associated	Valencia, Spain	454
Human Gut_1	4440452.7	54,632,274	229,857	human-associated	St. Louis, United States	454
Human Gut_2	4440616.3	174,824,393	507,928	human-associated	St. Louis, United States	454
Human Gut_3	4440611.3	103,097,122	526,727	human-associated	St. Louis, United States	454
Human Gut_4	4440826.3	124,768,172	499,499	human-associated	St. Louis, United States	454
Human Gut_5	4440824.3	100,520,072	414,497	human-associated	St. Louis, United States	454
Human Gut_6	4440639.3	93,430,618	440,521	human-associated	St. Louis, United States	454
Human Gut_7	4440461.5	105,923,024	522,134	human-associated	St. Louis, United States	454
Human Gut_8	4440613.3	102,979,597	312,665	human-associated	St. Louis, United States	454
Ocean_1	4443702.3	47,289,202	209,073	aquatic	Bergen, Norway	454
Ocean_2	4443707.3	31,359,337	135,033	aquatic	Bergen, Norway	454
Ocean_3	4443708.3	8,571,342	38,22	aquatic	Bergen, Norway	454
Ocean_4	4443703.3	30,991,689	134,915	aquatic	Bergen, Norway	454
Ocean_5	4443709.3	26,982,195	116,192	aquatic	Bergen, Norway	454
Ocean_6	4443705.3	68,187,679	304,02	aquatic	Bergen, Norway	454
Ocean_7	4443704.3	59,316,369	344,216	aquatic	Bergen, Norway	454
Ocean_8	4443706.3	38,021,523	162,871	aquatic	Bergen, Norway	454
Soil_1	4445996.3	116,821,792	312,444	soil	Navada, United States	454
Soil_2	4445993.3	133,555,260	352,417	soil	Navada, United States	454

Table S3.2 (cont.): Basic information summary of 48 publicly available metagenomic datasets from MG-RAST in five ecosystems (human, ocean, soil, freshwater and wastewater treatment plants) used in this study.

Label/Biomes	MG-RAST ID	Metagenome size (bp)	# Reads	Habitat Type	Location	Sequencing Method
Soil_3	4445994.3	254,548,462	683,082	soil	Navada, United States	454
Soil_4	4445990.3	219,117,356	583,724	soil	Navada, United States	454
Soil_5	4450750.3	87,160,647	239,933	soil	Navada, United States	454
Soil_6	4450752.3	76,860,743	233,279	soil	Navada, United States	454
Soil_7	4451103.3	397,257,248	1,040,697	soil	Navada, United States	454
Soil_8	4451104.3	347,578,191	998,484	soil	Navada, United States	454
Arctic Freshwater_1	4443683.3	101,310,476	100,085	freshwater	Antarctica	sanger
Arctic Freshwater_2	4443680.3	9,622,231	9,672	freshwater	Antarctica	sanger
Arctic Freshwater_3	4443679.3	9,755,315	10,042	freshwater	Antarctica	sanger
Arctic Freshwater_4	4443681.3	54,929,769	54,446	freshwater	Antarctica	sanger
Arctic Freshwater_5	4443682.3	284,069,722	283,663	freshwater	Antarctica	sanger
Arctic Freshwater_6	4443685.3	28,413,296	28,481	freshwater	Antarctica	sanger
Arctic Freshwater_7	4443686.3	101,573,008	103,058	freshwater	Antarctica	sanger
Arctic Freshwater_8	4443687.3	95,664,001	95,521	freshwater	Antarctica	sanger
Hydrocarbon WWTP_1	4507688.3	116,596,410	186,74	terrestrial	Inniskillen, Canada	454
Hydrocarbon WWTP_2	4507689.3	155,694,182	241,209	terrestrial	Inniskillen, Canada	454
Hydrocarbon WWTP_3	4509513.3	29,379,862	88,526	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_4	4509514.3	41,742,688	109,108	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_5	4509515.3	50,103,178	124,879	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_6	4509516.3	38,880,264	111,957	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_7	4523101.3	423,985,802	548,46	terrestrial	Fort McMurray, Canada	454
Hydrocarbon WWTP_8	4523103.3	340,253,804	436,518	terrestrial	Fort McMurray, Canada	454

Table S3.3: Relative abundance of microbial communities of the pre- and post-chlorinated samples from the wastewater treatment plants annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with that taxon divided by the total reads assigned for all domain.

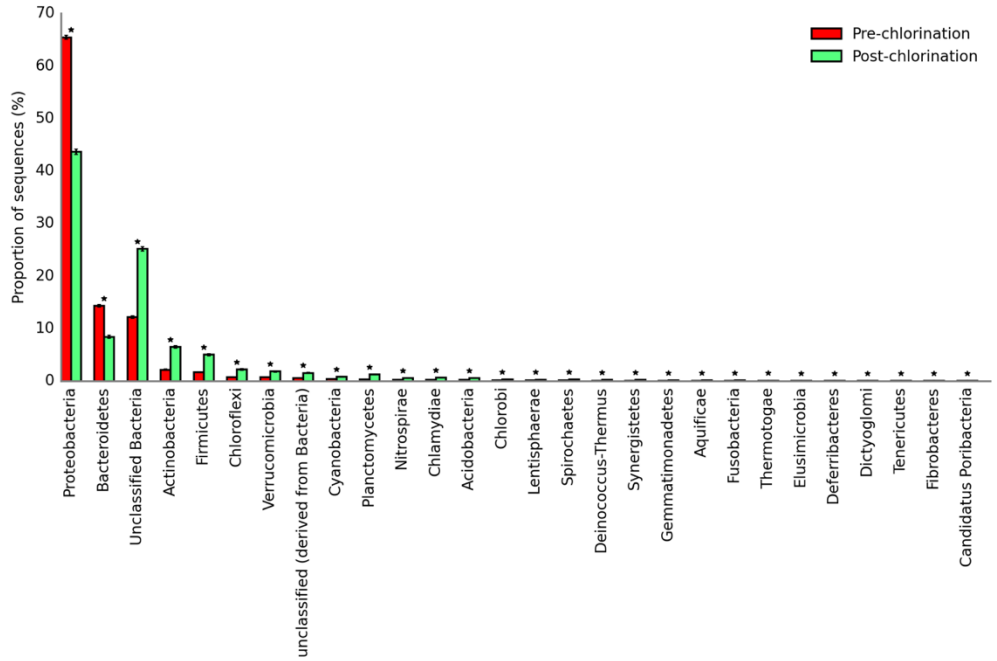
Domain	Plant A		Plant B	
	Pre-chlorination	Post-chlorination	Pre-chlorination	Post-chlorination
Bacteria	97,81	93,32	99,01	99,44
Eukaryota	1,20	4,37	0,65	0,24
Archaea	0,62	1,47	0,16	0,17
Viruses	0,18	0,37	0,08	0,03
Unclassified sequences	0,18	0,47	0,09	0,11
Other sequences	0,01	0,00	0,01	0,01



(**a**) (**b**)

Figure S3.1: Comparative taxonomic profile of the pre- and post-chlorinated wastewater in (**a**) Plant A and (**b**) Plant B at phylum level annotated by MG-RAST, assigned with the lowest common ancestor algorithm and statistically analysed with STAMP.

(**a**)



(**b**)

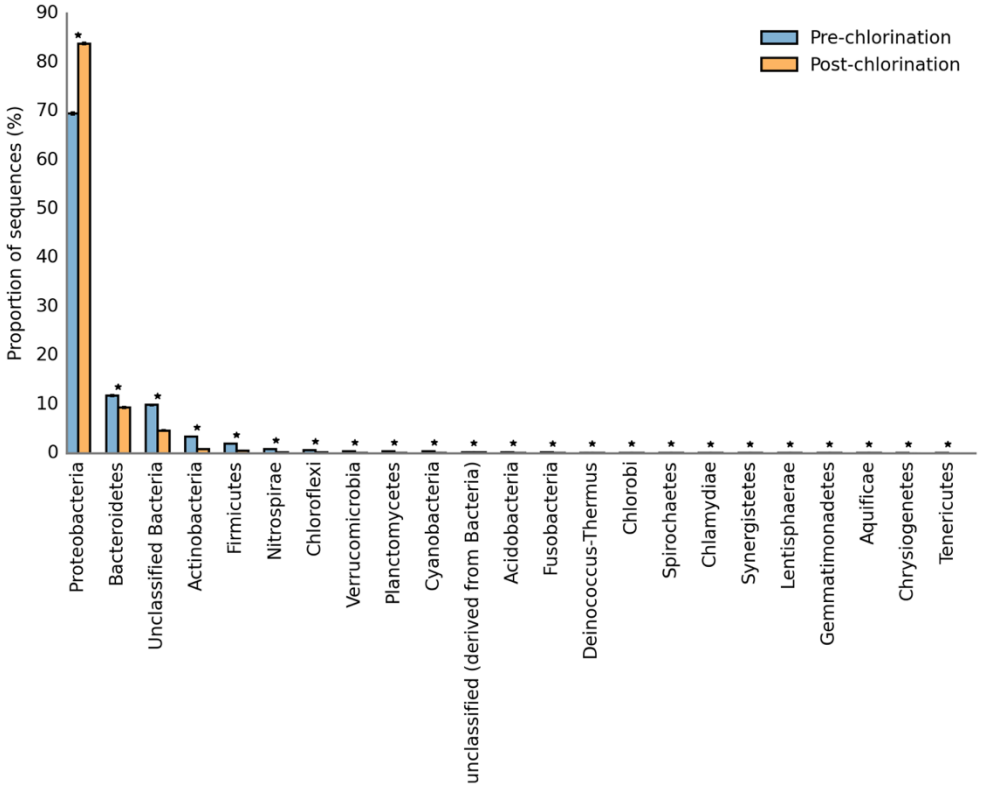


Figure S3.2: Comparative taxonomic profile of the pre- and post-chlorinated wastewater in (**a**) Plant A and (**b**) Plant B at phylum level annotated by MG-RAST and assigned with the lowest common ancestor algorithm. Asterisks indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP, whilst remaining phyla were omitted for clarity.

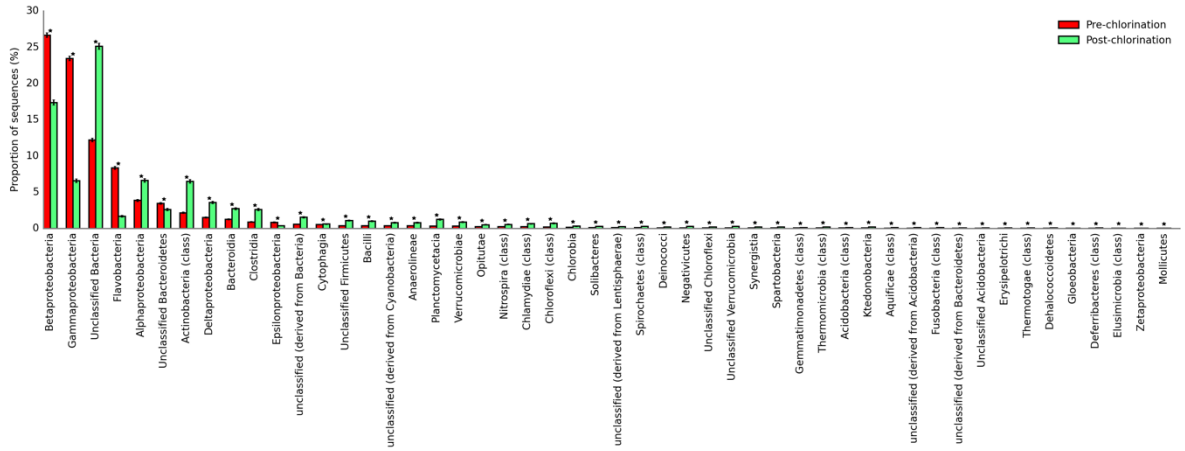


Figure S3.3: Comparative taxonomic profile of the pre- and post-chlorination disinfection wastewater in Plant A at class level annotated by MG-RAST and assigned with the lowest common ancestor algorithm. Asterisks indicate classes with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP, whilst remaining phyla were omitted for clarity.

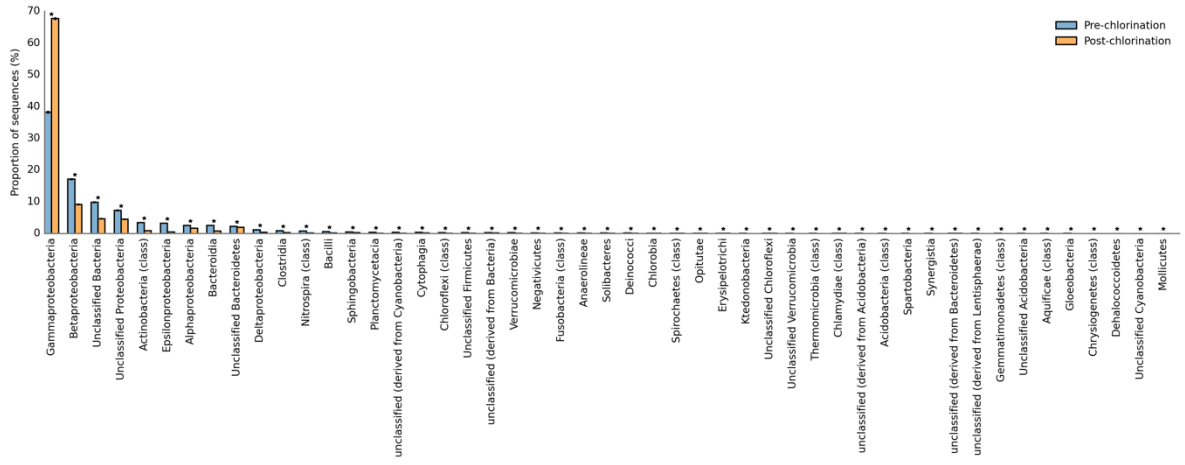


Figure S3.4: Comparative taxonomic profile of the pre- and post-chlorination disinfection wastewater in Plant B at class level annotated by MG-RAST and assigned with the lowest common ancestor algorithm. Asterisks indicate classes with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP, whilst remaining phyla were omitted for clarity.

Table S3.4: Relative abundance and distribution of predicted proteins assigned to level 1 SEED subsystems of the pre- and post-chlorinated samples for Plant A annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a subsystem divided by the total reads assigned in SEED subsystems.

SEED Level 1 Subsystems	Pre-chlorination relative abundance	Post-chlorination relative abundance
Amino Acids and Derivatives	8,23	7,25
Carbohydrates*	8,31	8,55
Cell Division and Cell Cycle*	1,57	1,73
Cell Wall and Capsule	3,85	3,78
Clustering-based subsystems*	15,51	16,30
Cofactors, Vitamins, Prosthetic Groups, Pigments	6,41	5,86
DNA Metabolism*	4,95	5,58
Dormancy and Sporulation	0,23	0,17
Fatty Acids, Lipids, and Isoprenoids	2,93	2,80
Iron acquisition and metabolism	0,84	0,49
Membrane Transport*	3,13	3,33
Metabolism of Aromatic Compounds	1,58	1,34
Miscellaneous	8,04	7,68
Motility and Chemotaxis	1,20	0,92
Nitrogen Metabolism	1,65	1,27
Nucleosides and Nucleotides*	2,81	2,85
Phages, Prophages, Transposable elements, Plasmids*	2,15	3,22
Phosphorus Metabolism	0,76	0,74
Photosynthesis*	0,11	0,11
Potassium metabolism	0,42	0,37
Protein Metabolism*	8,67	9,84
Regulation and Cell signalling**	1,63	4,63
Respiration	3,23	1,56
RNA Metabolism	4,60	3,16
Secondary Metabolism	0,35	0,32
Stress Response	2,40	2,25
Sulphur Metabolism	1,04	0,92
Virulence, Disease and Defence	3,41	3,00

*indicate a greater than 1-fold increase post-chlorination disinfection of the wastewater.

**indicate a greater than 2-fold increase post-chlorination disinfection of the wastewater.

Table S3.5: Relative abundance and distribution of predicted proteins assigned to level 1 SEED subsystems of the pre- and post-chlorinated samples for Plant B annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a subsystem divided by the total reads assigned in SEED subsystems.

SEED Level 1 Subsystems	Pre-chlorination relative abundance	Post-chlorination relative abundance
Amino Acids and Derivatives*	8,61	8,83
Carbohydrates	8,77	7,83
Cell Division and Cell Cycle*	1,56	1,75
Cell Wall and Capsule*	4,05	4,09
Clustering-based subsystems	15,25	14,97
Cofactors, Vitamins, Prosthetic Groups, Pigments*	6,48	6,85
DNA Metabolism	4,66	4,30
Dormancy and Sporulation	0,19	0,17
Fatty Acids, Lipids, and Isoprenoids*	3,04	3,11
Iron acquisition and metabolism*	0,91	1,26
Membrane Transport	3,19	3,14
Metabolism of Aromatic Compounds*	1,52	1,76
Miscellaneous*	8,14	8,22
Motility and Chemotaxis	1,06	0,57
Nitrogen Metabolism	1,53	1,25
Nucleosides and Nucleotides*	2,70	3,14
Phages, Prophages, Transposable elements, Plasmids	2,06	2,03
Phosphorus Metabolism*	0,86	0,92
Photosynthesis*	0,12	0,13
Potassium metabolism*	0,43	0,46
Protein Metabolism	8,41	8,20
Regulation and Cell signalling*	4,61	4,76
Respiration*	1,68	1,75
RNA Metabolism	3,10	3,06
Secondary Metabolism*	0,37	0,48
Stress Response	2,53	2,51
Sulphur Metabolism*	1,09	1,30
Virulence, Disease and Defence*	3,10	3,14

*indicate a greater than 1-fold increase post-chlorination disinfection of the wastewater.

**indicate a greater than 2-fold increase post-chlorination disinfection of the wastewater.

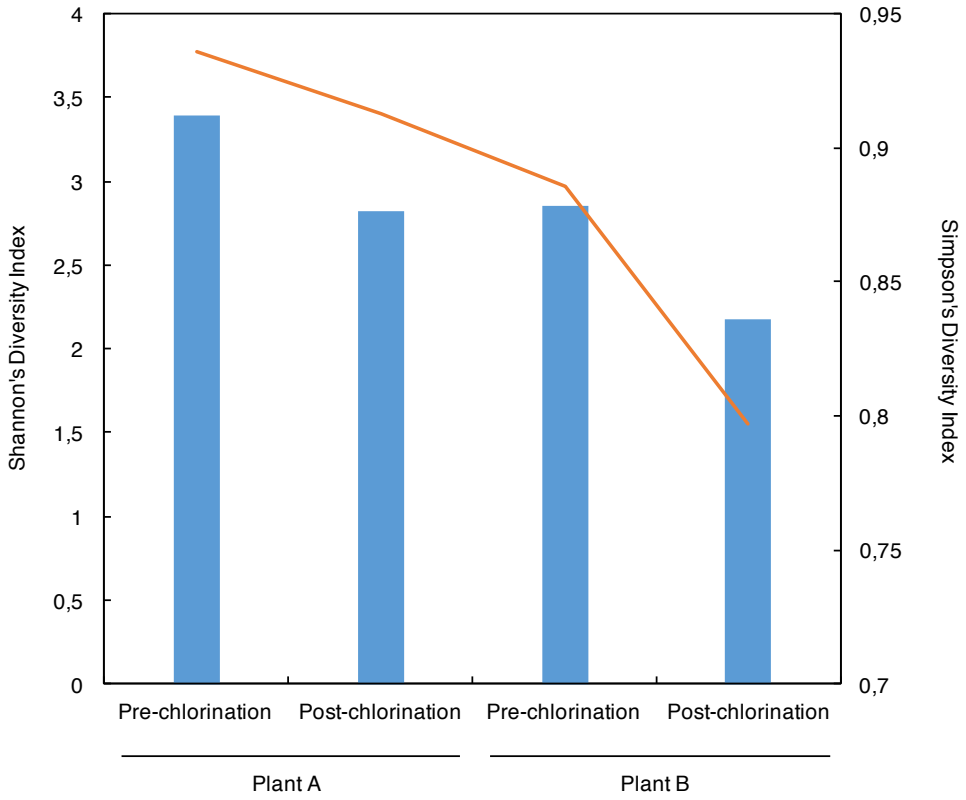


Figure S3.5: Shannon and Simpson diversity indices showing the diversity of antibiotic resistance genes (ARGs) of the bacterial communities pre- and post-chlorination at two different full-scale wastewater treatment plants. Indices were calculated in PAST software.

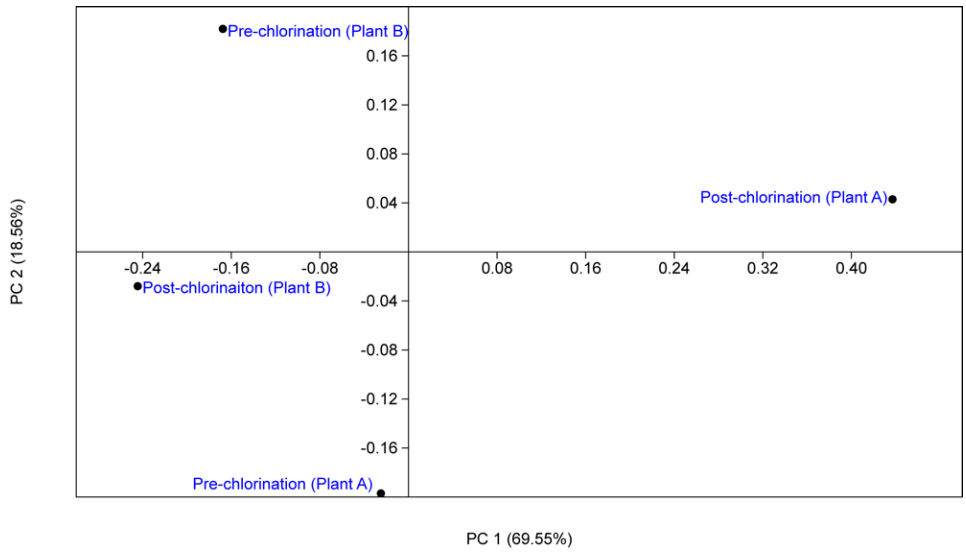


Figure S3.6: Principal coordinate analysis (PCoA) indicating the difference in antibiotic resistance genes (ARGs) profiles among all the samples tested in this study as determined using the PAST software.

Table S3.6: Diversity indices determined for ARGs profiles observed pre- and post-chlorination at Plant A and Plant B.

Indices	Plant A		Plant B	
	Pre-chlorination	**Post-chlorination**	**Pre-chlorination**	**Post-chlorination**
# Taxa	33	15	57	34
# Individuals	94	31	169	187
Simpson's Dominance	0,1145	0,2029	0,06418	0,08719
Simpson Diversity	0,8855	0,7971	0,9358	0,9128
Shannon Diversity	2,857	2,179	3,387	2,823
Buzas and Gibson's Evenness	0,5277	0,5895	0,5188	0,4948
Brillouin	2,448	1,702	2,979	2,577
Menhinick's Richness Index	3,404	2,694	4,385	2,486
Margalef's richness index	7,043	4,077	10,92	6,308
Equitability	0,8172	0,8048	0,8377	0,8005
Fisher's alpha	18,09	11,44	30,23	12,16
Berger-Parker	0,2979	0,4194	0,1716	0,1711
Chao 1	52,13	24	86	41,09

Chapter Four

Manuscript formatted and submitted to *Environmental Science and Pollution Research*

4 Antibiotic resistance genes, bacterial community structure and metabolic potential of two rivers impacted by treated wastewater effluent discharges as revealed by metagenomic analysis

4.1 Abstract

Treated effluents from most wastewater treatment plants (WWTPs) are routinely discharged into river bodies and have been implicated as major contributors to the dissemination of antibiotic resistance. Here we describe a metagenomic assessment of effluent discharge of two different WWTPs and their receiving rivers, in order to ascertain the impacts on the ecosystem functioning of the river. The antibiotic resistance genes, metabolic potential and bacterial community structures of the water samples were comprehensively investigated using whole metagenome shotgun pyrosequencing combined with metagenomic analysis. Overall, a shift in bacterial community composition of the river samples amongst members belonging to the Proteobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia and Firmicutes phyla was observed following effluent discharge. Similarly, a shift in the metabolic potential of the communities downstream of the effluent-receiving river compared to the upstream was observed. Alarmingly, diverse antibiotic resistance determinants, accounting for resistance to majority of clinically relevant classes of antibiotics, were recovered in all communities, further highlighting urban WWTPs as sources of antibiotic resistance. Results from this study shed more light on the role of WWTPs in the dissemination of antibiotic resistance into aquatic environments and the general impact of the discharge of improperly treated effluent on river ecosystems.

Keywords: Antibiotic resistance; Aquatic environment; Metagenomics; Microbial communities; River; Wastewater effluent

4.2 Introduction

Antibiotics have played a crucial role in the clinical management of most bacterial diseases since the 1940s. However, these drugs have been short lived and gradually rendered obsolete. A recent report has indicated an alarming global problem with regards to bacteria becoming resistant to the spectrum of commercially available antibiotics (WHO, 2014). This would mean that the standard treatment protocol currently adopted in clinical and animal sectors would no longer work, leading to a projected increase in mortality rates and risk of infection to others (WHO, 2014). Previous findings have already suggested the anthropogenic sources of antibiotic resistance genes (ARGs) with aquatic environments acting as major reservoirs (Marti et al., 2014). These findings highlight that ARGs can spread from environmental reservoirs to human pathogens. Specifically, urban wastewater treatment plants (WWTPs) have been implicated as a hotspot for horizontal transfer of ARGs between bacteria from difference origins due to the mixing of domestic, industrial, agricultural and clinical waste (Rizzo et al., 2013). Furthermore, subsequent dissemination of poorly treated wastewater effluents into aquatic sources have been shown to result in the spread of residual antibiotics (AB), antibiotic resistant bacteria (ARB) and ARGs into aquatic environments (Kümmerer, 2009; Lupo et al., 2012), providing conditions for horizontal exchange of ARGs and adding further selection pressure for ARB. The overall effect of WWTPs discharges on environmental resistome is poorly studied and has potentially far reaching influences on intensifying the antibiotic resistance problem.

Current microbiological tools used to assess the spread of ARGs and enrichment of ARB in these processes have either been culture- or molecular- dependent, each with its own advantages and limitations (Rizzo et al., 2013). Recently, the molecular approach has been

extended from identification and characterization of ARGs (Li et al., 2010; Luo et al., 2010; Storteboom et al., 2010) to genetic elements (Gaze et al., 2011; Johnning et al., 2013) found in cultured organisms. This allows for further understanding of the dynamics involved with capturing and transferring of ARGs and generation of ARB in microbial communities found in WWTPs. Additionally, whole-genome sequencing of isolated ARB has allowed for elucidation of mutation sites which give rise to ARGs as well as novel ARGs (Johnning et al., 2013). However, the use of culture-dependent methods to comprehensively characterize ARGs and ARB in complex microbial communities remains a challenge as it is a time consuming task. Furthermore as <1% of environmental organisms are culturable *in vitro* at laboratories with current technologies, determination of the complete spectrum of the environmental resistome (Schmieder and Edwards, 2012) is limited by this approach.

In the last few years, high-throughput sequencing combined with metagenomic analysis of a given sample has been considered a promising approach for the assessment of complex microbial communities (Kristiansson et al., 2011; Zhang et al., 2011; Wang et al., 2013). This approach has been applied for the assessment of various environmental scenarios, including soil (Howe et al., 2014) and river (Amos et al., 2014) microbiota, communities found in activated sludge from WWTPs (Ju et al., 2014), communities in sediments from an oil spill (Mason et al., 2014), faecal microbiota (Ilmberger et al., 2014), microbial diversity in aquatic settings (Doxey et al., 2014) as well as clinical settings such as the human gut profiling project (Qin et al., 2010), sputum analysis of patients with cystic fibrosis (Lim et al., 2014), viral pathogens detection (Yang, 2014) and many others. In determining the complex genetic composition and diversity present in aquatic samples with a relatively unbiased view (Gomez-Alvarez et al., 2009), a metagenomic insight is a favourable approach.

In this study, we investigated the impact of effluent discharge of two different urban WWTPs on the receiving river ecosystem. Specifically, the occurrence, abundance and diversity of ARGs, as well as the metabolic potential and structure of the bacterial communities were examined comprehensively. Results from this study may help to extend our knowledge on the contribution of effluent discharge from urban WWTPs in the dissemination of ARGs to river sources and its influence on the natural bacterial community metabolic potential and composition.

4.3 Materials and Methods

4.3.1 Site description and sample collections

In this study, two full-scale urban WWTPs, hereon designated as Plant A and Plant B, which discharges their final treated effluent (TE) directly into nearby rivers in the city of Durban, South Africa, were chosen for investigation. Plant A has a capacity of 70 megaliters/day with an operational capacity of 96% and uses the activated sludge and diffused air liquid technologies with gravity thickening, anaerobic digestion and belt press dewatering sludge technologies. The plant receives a mixture of nearby domestic and industrial wastewaters and discharges the final effluent into a nearby river in a suburban location. Receiving river samples were collected from approximately 500 m in opposite directions from the TE discharge point along the suburban river *viz.*, downstream (DS) and upstream (US). Plant B has a capacity of 25 megaliters/day with an operational capacity of 76% and uses the activated sludge liquid technology with anaerobic digestion and belt press dewatering sludge technologies. The plant receives a mixture of domestic, industrial, hospital wastewaters and discharges the final

effluent into a nearby river in a heavily urbanized location (Department of Water and Sanitation, 2015). Samples were also collected from the receiving river as described above along the urban river. All samples were collected in 5 L plastic bottles, pre-sterilized with 70% (v/v) ethanol and rinsed with 4 L of the sample at the various sampling sites prior to collection. Upon collection, the samples were transported on ice back to the laboratories within 3 h and stored at 4°C prior to DNA extraction which took place within 24 h.

4.3.2 Ethics statement

No special permits were required for this study. Permission for collection of wastewater

samples from both WWTPs was granted by the authorities of the respective WWTPs in Durban, South Africa.

4.3.3 Total DNA extraction and shotgun pyrosequencing

Prior to total DNA extraction, samples which visibly contained particles which could hinder DNA extraction was initially filtered through Whatman® filter paper #114 (Sigma-Aldrich, USA) to remove big particles and to allow for isolation of bacterial communities. Total DNA of the bacterial populations were extracted from the samples using PowerWater™ DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. This kit is specifically designed for isolating bacterial DNA from environmental water samples and includes inhibitor removal technology aimed at removing humic acid and other organic matter commonly found in environmental samples that can interfere with downstream analyses. The resulting purity and concentration of the DNA preparation was

determined using the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, USA) before samples were sent to the National Genomics Platform (Technology Innovation Agency, South Africa) for sequencing.

High-throughput whole metagenome shotgun pyrosequencing was conducted by the use of 454 GS FLX System (Roche, USA) with the GS FLX Titanium Rapid Library Preparation Kit (Roche, USA). The samples were barcoded by the use of GS FLX Titanium Rapid Library MID Adaptors Kit (Roche, USA) in order to enable multiplex sequencing. Five hundred micrograms of DNA was used for library construction. Briefly, DNA was nebulized using nitrogen gas and purified using MinElute PCR Purification Kit (QIAGEN, Germany) according to manufacturer's instructions. Fragment end repair and attachment of adaptors to the samples was carried out according to standard protocol (Roche, USA). Thereafter, the quality of the libraries was assessed by 2000 Bioanalyzer (Agilent Technologies, USA) using the High Sensitivity DNA Analysis Kit (Agilent Technologies, USA). Thereafter, relative fluorescence of the DNA libraries was measured on GloMax-Multi Microplate Multimode Reader (Promega, USA) using 6-fluorescein amidite (6-FAM) standards and 6-FAM-labeled Multiplex Identifiers (MIDs). The results were uploaded onto the Rapid Library Quantitation Calculator (Roche, USA) and rapid libraries were made with a final concentration of 1×10^7 molecules/µl. Emulsion PCR was carried out with the use of GS FLX Titanium MV emPCR Kit (Lib-L) (Roche, USA) according to manufacturer's instructions. Subsequently, four DNA copies per bead were enriched and used for the main sequencing run with *Escherichia coli* beads as a positive control (Roche, USA).

4.3.4 Data availability

All individual sequence reads have been deposited at the NCBI Short Read Archive (SRA) under the accessions no. SRR3629048, SRR3629049 and SRR3629050 for Plant A TE, US and DS river datasets, and SRR3629057, SRR3629055 and SRR3629056 for Plant B TE, US and DS river datasets.

4.3.5 Bioinformatic analysis

Raw data files from sequencing were de-multiplexed by removing the barcoded sequence and any secondary adapter sequences using in-house scripts. For initial quality control, raw reads were evaluated by CLC Genomics Workbench v.7.5 (CLC Bio, QIAGEN, Germany) quality control pipeline. In total, approximately 777,796 raw reads with an average length of of 613 bp corresponding to approximately 655.1 million bp were generated for this study. Specifically, raw sequences generated were 227,021, 115,483 and 169,225 reads for Plant A TE, DS and US of river receiving effluent from Plant A, respectively, and 180,557, 148,377 and 233,173 reads for Plant B TE, DS and US of the receiving river samples, respectively. The quality of all reads was within acceptable standards for the sequencing platform (see Table S4.1 in the supplemental material). Therefore, no reads were trimmed, filtered or discarded before implementation of the MG-RAST v.3.6 pipeline for further quality control, prediction and annotation (Meyer et al., 2008; Wilke et al., 2015). Ambiguous base filtering was first implemented by removing sequences with >5 ambiguous base pairs (bp) and length filtering by removing sequences with a length of >2 standard deviations from the mean were applied as quality control (QC) filters using SolexaQA (Cox et al., 2010) in the MG-RAST pipeline. Reads which passed the QC were used for further analysis. Upon ambiguous base filtering and

length filtering, approximately 9% (Plant A TE), 9% (DS), 7% (US), and 17% (Plant B TE), 9% (DS), 8% (US) of the reads were excluded from further analysis.

4.3.6 Combined functional and taxonomic annotations and assignments

Prior to functional annotation, reads within the datasets were screened for artificially replicated sequences and those identified sequences were removed using the dereplication tool (Gomez-Alvarez et al., 2009) implemented in the MG-RAST pipeline. Approximately 5% (TE), 5% (DS) and 6% (US) in Plant A, and 1% (TE), 5% (DS) and 6% (US) in Plant B of the reads were identified as artificially replicated sequences and removed (see Table S4.1 in the supplemental material). Thereafter, putative ORFs were identified using FragGeneScan, an *ab-initio* prokaryotic gene calling algorithm using the Hidden Markov Model (Rho et al., 2010), and their corresponding protein sequences were searched with BLAST against the M5nr protein database in the MG-RAST server (see Table S4.1 in the supplemental material). M5nr is a non-redundant protein database with integration of many public sequences databases including GenBank, SEED, IMG, UniProt, KEGG and eggNOGs (Meyer et al., 2008; Wilke et al., 2015). For functional assignments (i.e. assigned functions), analysis was performed using Hierarchical Classification approach against KEGG orthology (KO), SEED subsystems (SS) and clusters of orthologous groups (COG) with maximum e-value cut-off of 1×10^{-5}, minimum identity of 60% and a minimum alignment length of 15, which is measured in amino acids for protein databases and bp for RNA databases (Zheng et al., 2015).

For taxonomic identification and designation, an initial BLAT search against reduced RNA database (90% identity clustered version of SILVA database) was performed and the rRNA-

similar reads were then clustered at 97% identity with the longest read as the cluster representative. Thereafter, BLAT search of the cluster representative was performed against the M5rna database (see Table S4.1 in the supplemental material). M5rna is a ribosomal database which integrates SILVA, Greengenes and RDP databases (Meyer et al., 2008; Wilke et al., 2015). Taxonomic abundance was analyzed using the Lowest Common Ancestor (LCA) algorithm used in MEGAN (Huson et al., 2007) and implemented in MG-RAST with a maximum e-value cut-off, minimum identity and alignment length as described above. This algorithm assigns each read to the LCA from the set of matching taxa when BLASTx is applied. For example, if a given read had sequence similarity to 3 different families within the same order, the read is assigned at the order level rather than assigning the read to a specific family. Hence, the LCA algorithm has been reported to have lower rates of false positive assignments than the best hit classification algorithm implemented in the MG-RAST pipeline. However, this would result in a higher number of unspecific assignments or no hits in some cases (Huson et al., 2007).

Principle Coordinate Analysis (PCoA) of the Bray-Curtis dissimilarity coefficients based on the normalized abundance of functional assignments under the SS was used to identify the relationships between metabolic potential of the bacterial populations in this study and 48 publicly available metagenomes covering a wide variety of habitats (see Table S4.2 in the supplemental material). All the publicly available metagenomes utilized the same sequencing method as described for this study with the exception of Artic freshwater datasets which utilized Sanger sequencing. Annotation and functional assignment of the publicly available metagenomes was performed on the MG-RAST server to avoid potential bias from different analysis pipelines (Jeffries et al., 2011; Smith et al., 2012). Functional assignments were

performed using Hierarchical Classification approach against SS with maximum e-value cut-off, minimum identity and a minimum alignment length as described above. Prior to PCoA, SS assignments of each metagenomic datasets were normalized against the total number of hits in the database, log transformed and implemented in MG-RAST.

To determine the statistical differences between the metagenomic samples statistical analysis of metagenomic profiles (STAMP) v.2.1.3 (Parks and Beiko, 2010) software package was used. Statistical significance of differences between two samples (q values) was calculated on the basis of two-sided Fisher's exact test with Storey's false discovery rate (FDR). Because p values were not uniformly distributed using Storey's FDR, Benjamin-Hochberg FDR was applied for corrections during functional analysis. The confidence intervals were determined by Newcombe-Wilson's method. Statistical significance of differences between multiple samples was calculated with ANOVA analysis with a post-hoc test (Tukey-Kramer at 0.95) and an effect size (Eta-squared). A multiple-test correction using Benjamini-Hochberg FDR was employed. Only features with a q value of <0.05 were considered significant in this study.

4.3.7 Antibiotic resistance genes annotation and classifications

To identify ARGs in the metagenomes in order to establish the resistance profiles of the communities, the Comprehensive Antibiotic Resistance Database (CARD) was used for annotation (McArthur et al., 2013). The CARD database was preferred over other ARGs databases, such as Antibiotic Resistance Genes Online (Scaria et al., 2005), the microbial database of protein toxins, virulence factors, and antibiotic resistance genes (MvirDB) (Zhou et al., 2007), and Antibiotic Resistance Genes Database (Liu and Pop, 2009), because these

databases are neither exhaustive nor regularly updated (Gupta et al., 2014). BLASTx against the database was carried out using DIAMOND tool v.0.7.11 (Buchfink et al., 2014) with an e-value cut-off of 1×10^{-5}. A read with its best BLAST hit was deemed ARG-like if the hit had a sequence similarity of >90% over an alignment of ≥25 amino acids (Chao et al., 2013; Wang et al., 2013). Although such a high similarity threshold excluded some divergent ARGs from the analysis, we still used a more conservative strategy and thus only focus on those that are highly similar to the known ARGs. The classification of ARG-like sequences was performed using the structured database of CARD (McArthur et al., 2013) and in-house written scripts. Visualization of the distributions of ARGs classes and their abundances in the total annotated ARGs from the metagenomes was constructed and visualized using Circos v.0.69.2 (Krzywinski et al., 2009). Using the PAleontological Statistics (PAST) software v.3 (Hammer et al., 2001), Mann-Whitney test was implemented to compare whether ARGs abundances were significantly different among the various sample environments (Hu et al., 2013; Li et al., 2015). Various diversity indices and PCoA based on the Bray-Curtis dissimilarity coefficients of the abundance of ARGs were determined using the 'vegan' package (Shen and Fulthorpe, 2015) from the program R (R Development Core Team, 2007) with a bootstrap of 1000.

4.4 Results

4.4.1 Taxonomic profile of microbial communities

Taxonomic assignment of both predicted rRNA genes sequences and predicted proteins in all metagenomic datasets were conducted based on all the available annotation source databases in MG-RAST.

Domain distributions in the six water samples, showed the expected dominance of bacteria across all metagenomes accounting for 93.32%, 98.37% and 98.94% of the metagenomes for Plant A TE, DS and US of the receiving suburban river samples, respectively, and 99.44%, 99.52% and 98.69% for Plant B TE, DS and US of the receiving urban river samples, respectively. The eukaryotic domain was observed to be dominated by members of the Oligohymenophorea class in TE sample of Plant A, whilst members of the Coscinodiscophyceae class were dominant in both the DS and US suburban river samples. In contrast, potentially novel members of the Streptophyta class (determined as unclassified) were observed to dominate the eukaryotic domain in Plant B TE and in both DS and US urban river samples. The remaining sequences were distributed amongst archeal and viral domains in very low abundance (see Table S4.3 in the supplemental material).

For a better understanding of the microbial community structure in all the metagenomes, taxonomic affiliation at different levels was analysed. At the phyla taxonomic level, bacterial communities were diverse with representatives of 27 phyla and 1 novel bacterial phyla (determined as unclassified, however, under the Bacterial domain level) found across all samples. Proteobacteria was observed as the dominant phylum in all samples accounting for 43.68% to 80.07% of the sequences in Plant A TE and discharged effluent-receiving river samples, and 60.84% to 83.81% of the sequences in Plant B TE and discharged effluent-receiving river samples (Figure 4.1). Following Proteobacteria, dominance of Bacteroidetes, Actinobacteria and Firmicutes phyla was observed in Plant A TE. However, a shift to Bacteroidetes, Actinobacteria and Verrucomicrobia phyla was noted DS of the river receiving effluent from Plant A. The same members of the phyla were also observed US of the river receiving effluent discharged by Plant A. In the case of Plant B, following Proteobacteria,

dominance by Bacteroidetes, Actinobacteria and Verrucomicrobia was observed in US river samples. However, both TE and DS samples saw Bacteroidetes, Actinobacteria and Firmicutes as the dominant phyla. Statistical analysis using STAMP indicated that several significant differences (q<0.05) were observed at this taxonomic level (see Figure S4.1 in the supplemental material). Notably, significant decreases (q<0.05) in Proteobacteria members were observed DS compared to US communities along both rivers, whilst significant increases (q<0.05) in Bacteroidetes, Actinobacteria and Firmicutes were observed in DS communities compared to US communities along both rivers.

At the class taxonomic level, the bacterial community was diverse with representatives of more than 50 classes observed across all samples. Beta-proteobacteria, Gamma-proteobacteria, Flavobacteria and Alpha-proteobacteria were the most abundant classes in treated effluents of both Plants as well as in the microbial communities DS and US of the receiving rivers, accounting for 1.63-67.69%, 2.08-57.93% and 3.63-51.94% of the assigned reads, respectively (Figure 4.2). Although Beta-proteobacteria dominated the community, significant increases (q<0.05) were observed in bacterial communities DS of the river receiving effluent discharge from Plant A compared to US river communities. Conversely, significant decreases (q<0.05) were observed for Gamma-proteobacteria, Flavobacteria and Alpha-proteobacteria DS of the river receiving effluent discharge from Plant A compared to US river communities. In Plant B, Gamma-proteobacteria dominated the microbial community DS of the effluent-receiving river in significant (q<0.05) abundance relative to US river communities, whilst significant decreases (q<0.05) were seen in Betaproteobacteria, Flavobacteria and Alphaproteobacteria classes in DS river communities (see Figure S4.2 in the supplemental material).

4.4.2 Composition, abundance and diversity of antibiotic resistance genes

From the metagenome of the effluent discharged by Plant A, a total of 31 reads were identified as ARGs. However, in the DS and US sample of the receiving river, a total of 43 and 68 reads were identified as ARGs, respectively. In contrast, a total of 187 reads were annotated as ARGs in the metagenome of effluent discharged by Plant B. However, a 82.5% increase in the number of reads assigned as ARGs was observed from US (57) to the DS (104) of the effluent-receiving river. There was no significant difference in the total abundance of the ARGs detected at US and DS of both rivers separately receiving the treated effluent from each Plant.

Among the 27 ARGs types in the structured CARD database, a total of 13 were detected in Plant A TE, DS and US of the effluent-receiving river samples, with 9 types shared among the three metagenomic samples. For Plant B, a total of 14 types were detected in TE, DS and US of the receiving river samples, with 9 types also shared among the metagenomic samples. The different types of ARGs conferring resistance to most major classes of antibiotics were noted to be distributed and relatively abundant across all samples in this study (Figure 4.3). Calculated diversity indices consistently indicated a reduced diversity of ARGs types in DS communities of river receiving effluent from Plant A compared to US sample, although, it should be noted that the ARGs diversity in the TE was the highest amongst all three samples. Conversely, an augmented diversity of ARGs types was observed in DS sample of river receiving Plant B TE compared to US diversity. It is worth noting that the ARGs diversity in Plant B TE was also the highest amongst all three samples (see Figure S4.3 in the supplemental material). Furthermore, PCoA of the ARGs class profiles showed separate clustering between the various samples with different environmental sources (see Figure S4.4 in the supplemental material).

The elfamycin class of ARGs was the dominant class in TE sample, and DS and US of the river receiving effluent from Plant A, accounting for 41.93-51.16% of the relative abundance, followed by resistance genes for fluoroquinolones (2.94-13.05%), rifampins (4.41-12.9%), sulfonamides (9.3-10.29%), multidrug resistance genes (3.23-10.29%) and aminoglycosides (3.23-7.35%) (Figure 4.3). However, distribution of the remaining ARGs classes showed relatively the same Buzas and Gibson's evenness index ranging from 0.5257 to 0.6537. The elfamycin class of ARGs was also the dominant class in the US communities of Plant B TE receiving river, accounting for 43.86% of determined ARGs in the population. However, aminocoumarin class of ARGs was found to be dominant in both the Plant B TE and DS of the receiving river , accounting for 17.65% and 20.04%, respectively. This was followed by resistance genes for tetracycline (1.76-14.34%), multidrug resistance genes (7.02-9.63%) and rifampins (4.81-17.54%). Distribution of the various resistance class were not even as indicated by the Buzas and Gibson's evenness index, with a lower index obtained for the US sample of Plant B effluent-receiving river communities compared to Plant B TE and DS river sample (see Figure S4.3 in the supplemental material).

Under the elfamycin ARG class, variants of the elongation factor Tu were the dominant type across all samples. Point mutations that occurs in *Mycobacterium tuberculosis* and *Escherichia coli* beta-subunit of RNA polymerase (*rpoB*) conferring resistance to rifampicin was found to be in high abundance in Plant A TE and DS of the effluent discharge receiving river. Additionally, genes such as *TEM-1*, *tetQ*, *sul2* and *tet39* were also detected in these samples (see Figure S4.5 in the supplemental material). Conversely, genes such as *adeJ, msrE, mexT, OXA-211* were found to be in high abundance in Plant B TE and DS of the TE-receiving river. Furthermore, point mutation that occurs in *M. tuberculosis*, *E. coli* and *Staphylococcus aureus*

beta-subunit of RNA polymerase (*rpoB*) and DNA subunit B gyrase (*gyrB*) were also in abundance in these samples (see Figure S4.5 in the supplemental material).

4.4.3 Functional analysis using SEED, KEGG and COG identifiers

Overall, functional analysis of the metagenomes showed approximately 36% to 95% of the predicted open reading frames (ORFs) to be assigned to the SS functional levels in Plant A TE, DS and US metagenomes, whilst approximately 79% to 86% of the predicted ORFs were assigned to the SS functional levels in Plant B TE, DS and US samples. For KEGG pathways, approximately 11% to 28% of the predicted ORFs were for Plant A TE, DS and US metagenomes, whilst approximately 23% to 26% of the predicted ORFs were assigned for Plant B TE, DS and US metagenomes. Similarly, approximately 18% to 47% of the predicted ORFs were assigned by COG for Plant A TE, DS and US metagenomes, whilst approximately 29% to 44% of the predicted ORFs were assigned to the SS functional levels for Plant B TE, DS and US of the river communities (see Table S4.1 in the supplemental material).

At the level 1 of SEED classifications, in Plant A TE, DS and US of the receiving river metagenomes, approximately 9-16% of the genes were related to the subsystems of clustering and protein metabolism whereas, 8-9% of the sequences were related to subsystems of carbohydrate and miscellaneous. About 6-8% of the genes were related to subsystems of amino acids and derivatives, and cofactors, vitamins, prosthetic groups, and pigments. While 3-6% of the genes were related to subsystems of DNA metabolism, RNA metabolism, and membrane transport. A relatively low number of the genes were related to the subsystems of virulence (3%) in all the metagenomes, whereas 5%, 2%, 1% of the predicted protein sequences were

related to the subsystems of cell wall and capsule, stress response, and motility and chemotaxis, respectively. An additional 2% of the genes were related to subsystems of regulation and cell signalling (Figure 4.4a and Table S4.4 in the supplemental material). Furthermore, SS statistical analysis showed 4 out of 28 level 1 subsystems were significantly different (q<0.05) between DS and US metagenomes. Specifically, significant decrease (q<0.05) in carbohydrate and phages, prophages, transposable elements and plasmids subsystems were observed in DS compared to US communities. However, significant increase (q<0.05) in cell wall and capsule, and nitrogen metabolism subsystems were observed in DS compared to US communities (see Table S4.4 in the supplemental material). Statistical analysis of functional level subsystems assignments between DS and the US river metagenome in Plant A showed slight similarities (r^2=0.89) between the metabolic potential of the communities (see Figure S4.6a in the supplemental material). This was further observed in principal component analysis (PCA) which clustered DS and US river metagenomes closer to each other than with TE metagenome (see Figure S4.7a in the supplemental material).

In the metagenomes of Plant B TE, DS and US of the receiving river, approximately 9-15% of the genes were related to the subsystems of clustering and amino acids and derivatives, whereas, approximately 8% of the sequences were related to both subsystems of miscellaneous and protein metabolism. About 6-10% of the genes were related to subsystems of carbohydrates, and cofactors, vitamins, prosthetic groups, and pigments, respectively, while 4% and 5% of the genes were related to subsystems of DNA metabolism and RNA metabolism, respectively. A relatively low number of the genes were also related to the subsystems of virulence (3%) in all the metagenomes, whereas 4%, 3%, 2% of the predicted protein sequences were related to the subsystems of cell wall and capsule, stress response, and motility and chemotaxis, respectively.

An additional 2% of the genes were related to subsystems of regulation and cell signalling (Figure 4.4a and see Table S4.5 in the supplemental material). Furthermore, SS statistical analysis showed 16 out of 28 level 1 subsystems to be significantly different (q<0.05) between the DS and US metagenomes. Specifically, significant decrease (q<0.05) in carbohydrate, miscellaneous, RNA metabolism, membrane transport, motility and chemotaxis subsystems were observed in DS compared to US communities. However, significant increase (q<0.05) in clustering, amino acids and derivatives, fatty acids, lipids, isoprenoids, DNA metabolism, nucleosides, nucleotides, metabolism of aromatic compounds, phages, prophages, transposable elements, plasmids, cell division and cell cycle, phosphorus metabolism, sulphur metabolism and secondary metabolism subsystems were observed in DS compared to US communities (see Table S4.5 in the supplemental material). Statistical analysis of the functional level subsystems assignments between DS and the US river metagenome in Plant B revealed no similarities (r^2=0.791) between the metabolic potential of the communities. (see Figure S4.6b in the supplemental material). This was further observed in principal component analysis (PCA) which clustered DS and TE metagenomes closer to each other than with US metagenome (see Figure S4.7b in the supplemental material).

The KEGG mapping and subsequent assignment of the predicted protein sequences identified 1881, 1721 and 1836 unique genes associated with KO identifiers from the Plant A TE, DS and US of the receiving river metagenomes, respectively. Majority of the predicted protein sequences were associated with the functions relating to metabolism (54-60%), followed by genetic information processing (19-24%) and environmental information processing (14-15%) (Figure 4.5a). Statistical analysis indicated only three pathways (lipopolysaccharide biosynthesis, ABC transporters and flagellar assembly) were significantly different (q<0.05)

between DS and US metagenomes of Plant A effluent-receiving river at the level 3 SS (see Figure S4.8a in the supplemental material). For Plant B TE, DS and US of the receiving river metagenomes, mapping and assignment identified 1607, 1668 and 1982 unique genes associated with KO identifiers, respectively. Majority of the predicted protein sequences were associated with the functions relating to metabolism (38-60%), followed by genetic information processing (12-18%) and environmental information processing (9-15%). It must be noted that a large proportion (12-37%) of the identified KO genes were observed to be uncategorized for the DS and US samples of this river metagenomes (Figure 4.5a). Statistical analysis indicated 39 pathways which were significantly different (q<0.05) between DS and US metagenomes of Plant B TE-receiving river at the level 3 SS (see Figure S4.8b in the supplemental material).

COG families revealed significant differences (q<0.05) in the overall cellular processes, signalling and metabolism between DS and US river metagenomes in Plant A (see Figure S4.9a in the supplemental material). Furthermore, statistical comparison of both samples to Plant A TE showed a closer relationship between the US and DS than to TE metagenome (see Figure S4.10a in the supplemental material). Conversely, all families were observed to be significantly different (q<0.05) between DS and US of the metagenomes of Plant B TE-receiving river (see Figure S4.9b in the supplemental material). This was also observed in the statistical comparison of both samples to Plant B TE, which showed a closer relationship between DS and TE metagenomes than to US metagenome (see Figure S4.10b in the supplemental material). Overall, general function predictions (e.g. ABC-type transport systems, permeases, ATPases, hydrolases, peptidases, phosphatases), functions relating to amino acid metabolism and transport, energy production and conversion, translation systems, ribosomal biogenesis,

replication systems, recombination (transposase) and repair systems, cell wall or membrane biogenesis, families implicated in signalling processes (e.g. EAL domain, GGDEF domain, signal transducing histidine kinase, CheY-like receiver) were present predominantly in all river samples. It should be noted that functions relating to inorganic ion transport and metabolism (e.g. arsenite and silver efflux pumps, cation transport ATPase, Fe, Mn transporters) were found in higher prevalence in the US and DS river receiving Plant B TE compared to the river receiving effluent discharges from Plant A (Figure 4.5b).

To extend comparative metagenomic analyses, comparison of the overall metabolic potential of the bacterial communities of all metagenomic samples from this study to 6 different environmental habitats by principal coordinate analysis (PCoA) revealed that different ecosystems clustered differently (Figure 4.6). For example, human oral, human gut, ocean and soil metagenomes cluster into individual groups. Conversely, Arctic freshwater and hydrocarbon wastewater samples showed variations in metagenomes although sampled from the same ecosystem. Samples from this study were observed to ordinate around ocean, soil and Artic freshwater metagenomes.

4.5 Discussion

Although antibiotic resistance has become a major threat to human health worldwide, this phenomenon has been overlooked in aquatic settings. Aquatic sources may provide an ideal setting for the acquisition and dissemination of antibiotic resistance because they are frequently

impacted by anthropogenic activities, such as effluent discharges from urban WWTP. In this study, whole metagenome shotgun pyrosequencing combined with metagenomic analysis was used to explore the profiles of antibiotic resistance genes, metabolic potential and bacterial community structure in two rivers impacted by effluent discharges from separate WWTP. Sequencing of the metagenomic libraries has allowed for a deeper insight into the complex communities in the effluents discharged by these WWTPs and their receiving rivers, with thousands of reads generated and assigned to different taxa, metabolic and ARG categories.

4.5.1 Phylogenetic signature of metagenomic sequences

Proteobacteria was observed to be the dominant phyla across all samples analysed in this study. This finding corroborates several studies investigating different ecosystems, such as: those targeting the 16S rRNA gene of samples collected from WWTPs (Miura et al., 2007; Silva et al., 2010), DNA cloning of activated sludge (Snaidr et al., 1997), microarray analysis of biological wastewater treatment reactors (Xia et al., 2010), metagenomic analysis of different WWTPs (Hu et al., 2012), drinking water treatment plants (Shi et al., 2013), artic river sources (Kolmakova et al., 2014), soil (Roesch et al., 2007) and freshwater sources (Ghai et al., 2011). Therefore, predominance of members belonging to this phylum in such environments could be explained by the fact that the phylum comprises of one of the most phylogenetically and metabolically versatile group in the Bacteria domain (Ettema and Andersson, 2009). Furthermore, a study which examined the global patterns of bacterial communities from different habitats suggested that Proteobacteria typically occupies an average of 40% of a bacterial population (Nemergut et al., 2011). Hence, as WWTPs and rivers present with a constant changing environment, such as the composition and concentration of nutrients and/or pollutants, the observation dominance by this phylum across all samples is not surprising.

Besides dominance by Proteobacteria, high abundance of several other phyla was seen downstream of both rivers in this study, including members belonging to the Bacteroidetes, Actinobacteria, Verrucomicrobia and Firmicutes phyla. Detection of Bacteroidetes, Actinobacteria and Firmicutes has been well documented in recent studies investigating WWTPs, such as untreated wastewaters (Shanks et al., 2013), swine WWTP (Da Silva et al., 2015), anaerobic reactor digesting activated sludge from WWTP (Guo et al., 2015), tannery WWTP (Wang et al., 2013) and activated sludge of a WWTP in Hong Kong (Yu and Zhang, 2012). Additionally, members of these phyla have also been reported in studies investigating river sources (Sánchez-Andrea et al., 2011; Wu et al., 2012; Jordaan and Bezuidenhout, 2013; Kolmakova et al., 2014), lake sediments (Sauvain et al., 2013), mangrove (Andreote et al., 2012) and soil (Foong et al., 2010; Zhang et al., 2015) sources. Although, members of the Verrucomicrobia phylum have not been widely reported in WWTP environments, studies have indicated dominance by this phyla in aquatic sources (Brown et al., 2015; Satinsky et al., 2015). Furthermore, it has been suggested that the core human microbial signature is composed of Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria (Buffie and Pamer, 2013). Therefore, presence and abundance of the observed phyla in the receiving rivers is not surprising and suggests that at this taxonomic level of complexity, effluents discharge from these WWTPs do not disrupt the natural bacterial community structure in the receiving river sources.

Phylogenetic analysis at the class taxonomic level revealed members belonging to the Beta-proteobacteria, Gamma-proteobacteria, Flavobacteria and Alpha-proteobacteria classes to be in high abundance. This result is consistent with several studies investigating different aquatic environments (Berg et al., 2009; Mlejnková and Sovová, 2010; Liu et al., 2012; Du et al., 2013;

Chiaramonte et al., 2014). However, fluctuations in the community structure at this level was apparent when upstream and downstream samples were compared to each other, suggesting the shifts in community structure as a consequence of the WWTP effluent discharged. This is in line with many cases where wastewater effluent discharges make up a significant proportion of the receiving water body (Brooks et al. 2006; Gücker et al. 2006). Equally, inadequately treated wastewater effluents have been said to be sources of organic and inorganic nutrients, heavy metals, pesticides, pharmaceuticals and antibiotics (Drury et al., 2013; Proia et al., 2013). Therefore, their presence may lead to beneficial or detrimental effects on the microbial communities and changes in the community structure and diversity as a result of the effluent discharge is expected on a case to case basis. Previous findings have suggested an increase (Garnier et al., 1992; Wakelin et al., 2008) and decrease (Drury et al., 2013) in bacterial community diversity and composition from rivers receiving WWTP discharges. Thus, these reasons could explain the observed shift between genera belonging to the Betaproteobacteria, Gammaproteobacteria, Flavobacteria and Alphaproteobacteria classes from upstream and downstream sites along both rivers.

4.5.2 Metabolic potential of treated wastewater effluent discharge and receiving rivers

Whole genome shotgun sequencing of metagenomes has the capacity to fully sequence the majority of available genomes within an environmental sample (or community). Therefore, besides community composition profiles, metabolic potential of microbial communities can be explored. As with the diverse bacterial communities observed in this study, a wide metabolic potential was also present across the metagenomic samples. The percentage of annotated and successfully assigned reads to either SS, KO or COG identifiers in this study were higher than those previously reported using the Illumina sequencing platform used to characterise structure

and function of a sewage sludge community (Yu and Zhang, 2012). However, previous studies have reported that approximately 25% of Illumina reads and >36% of pyrosequencing reads from soil metagenomes had a significant match to the SS (Uroz et al., 2013). Furthermore, functional metagenomic investigations of microbial communities in a hydrothermal systems using pyrosequencing (Tang et al., 2013) showed comparable results to this study. Hence, divergences of annotation proportions may perhaps be due to the difference in environmental sample types, community composition and sequencing platforms in those studies.

As expected, genes associated with the metabolism of carbohydrates, protein metabolism, amino acids systems were observed to be in relatively high abundance as they are related to the housekeeping functions of all living organisms. In addition, clustering-based subsystems (containing functions such as proteosomes, ribosomes, and recombination-related clusters) and miscellaneous subsystems (containing genes associated with iron-sulphur cluster assembly, common prokaryotic-plant genes and Niacine-Choline transport and metabolism) were also found to be in relatively high abundance. These findings are in agreement with previous studies investigating sewage sludge (Yu and Zhang, 2012), activated sludge (Dinsdale et al., 2008), soils (Manoharan et al., 2015) and tropical freshwater sediment (Costa et al., 2015), which presents with similar environmental conditions as the samples investigated in this study. Therefore, this suggests that genes associated with these functional categories could form the core metabolic potential for bacterial communities in wastewater and receiving river environments. However, further omic-based investigations should be conducted to further confirm this assumption.

Interestingly, distinct differences were observed between the metabolic potential of the upstream compared to the downstream river samples only in rivers receiving treated effluent from Plant B. This suggests that Plant B WWTP effluent discharges could have contributed to the shift in the metabolic potential of the communities present in the receiving river. This also illustrates the concept of biotic homogenization, which suggests that "anthropogenic modifications of the environment are reducing the biological richness that exist in natural ecosystems". Consequently only a subset of naturally occurring species are constantly being selected for as a result of the human-altered ecosystems (McKinney 2008; Baiser et al. 2012). Although, this phenomenon has been largely seen in plant and animal community studies, it has been poorly documented for microbial communities and their metabolic potential (McKinney, 2008). Effects of this phenomenon has been suggested to result in a more homogenized biosphere with lower diversity on regional and global scales (McKinney and Lockwood, 1999). Therefore, our results suggests that Plant B WWTP effluent may be a driving force for biotic homogenization of bacterial communities in this receiving river and may lead to negative implications on the ecosystem functions. These results were also observed utilizing T-RFLP fingerprinting technique (see Chapter Five).

Overall, comparisons of the metabolic potential of the metagenomes of samples in this study to those from other ecological settings revealed considerable functional differences between unique ecosystems. Though, different ecosystems displayed distinct characteristics, metagenomic samples from this study shared similar characteristics with ocean and soil ecosystems. These observations are expected as WWTPs and rivers present with an environment closer to soil and ocean environments than that of human settings.

4.5.3 Proportion of putative antibiotic resistance gene sequences differs across metagenomes

The strict and unambiguous antibiotic resistance ontology of the CARD database was used to annotate genes as ARG-like in this study. It should be noted that the diversity and abundance of the antibiotic resistome observed by CARD alignment were different from those derived from the MG-RAST analysis of the metagenome. Possible reasons for the divergence may be the difference in the reference database used as well as the contribution of the alignment algorithm. MG-RAST relies on the use of BLAT algorithm for similarity searches, which is less sensitive than the BLAST algorithm (Yu and Zhang, 2013). Additionally, the CARD database is a highly curated and comprehensive database (McArthur et al., 2013) compared to the incomplete resistance to antibiotics and toxic compounds subsystem on ARGs (Wang et al., 2013).

Genes conferring resistance to the elfamycin and aminocoumarin class of antibiotics was observed to be the most abundant across all metagenomes. The aminocoumarin family of antibiotics inhibits the beta sub-unit of DNA gyrase which is essential in cell division (Galm et al., 2004). The elfamycin family of antibiotics inhibits protein synthesis of Gram-negative bacteria by interacting with elongation factor Tu (Wolf et al., 1974; Hall et al., 1989; Vogeley et al., 2001). Specifically, majority of the elfamycin-like ARGs in the metagenomes were associated with the variants of the elongation factor Tu. Elfamycin are not commonly used in treatment of human diseases due to the very narrow spectrum against human pathogens (Miele et al., 1994), however, it is applied as a growth promoting agent in animals owing to its excellent activities (Brötz et al., 2011; Maiese et al., 1989). Though, aminocoumarin is used in the treatment of human diseases, it is also largely applied for treatment of *Stapholococcus*

infections, cholera infections and treatment of bovine mastitis in animal food production (Economou and Gousia, 2015; Marshall and Levy, 2011). Therefore, high abundance of these ARGs in the metagenomes suggests that the major source of wastewaters received by both WWTPs could be from nearby animal feed industries. This reason could also partially explain the coinciding high occurrence of genes conferring resistance to tetracycline class of antibiotics as tetracycline is commonly used for livestock purposes in South Africa (Eagar et al., 2012). Furthermore, studies suggests that sources of *tet39* gene could result from clinical (Agersø and Guardabassi, 2005) or animal sources (Agersø and Petersen, 2007) and are common amongst *Acinetobacter* spp. (Hamidian et al., 2016), which is a cause for concern with the increasing threat in hospitals from multidrug-resistant *Acinetobacter* strains (Dijkshoorn et al., 2007). The occurrence of these particular family of antibiotic and their ineffective biodegradation in WWTP may allow for low concentrations to persist in treated effluents leading to the selection of these ARGs (Gullberg et al., 2011) and their subsequent dissemination to the effluent receiving rivers.

Surprisingly, upstream and downstream samples contained ARGs without salient differences between each another despite the WWTP effluent discharge. However, slight differences were by all means detected across all the samples, indicating that the ARGs load from the WWTP effluent discharges did not have a detectable influence on the receiving rivers. This result is in contradiction to the fact that WWTPs represent the main sites through which ARGs are released into the environment (Marti et al., 2014). However, a recent study by Munck et al. (2015) suggested that the abundant WWTP resistome is only disseminated, to a very limited extent, to other microbial communities in differing environments. Furthermore, the study suggested that members of the WWTP core resistome rarely take part in gene exchange networks with human

pathogens with mobilization acting as the main barrier preventing the spread of ARGs (Munck et al., 2015). Hence, the findings of this study suggest that the WWTP effluents discharged into receiving rivers did not have major influences on ARGs dissemination into the aquatic sources.

Nonetheless, the proportions of the ARGs identified in this study were diverse and comparable to the results from previous studies investigating activated sludge (Zhang et al., 2011), sewage effluent (Port et al., 2012), plasmids recovered from WWTP (Li et al., 2015), non-hospital medical care facility (Bäumlisberger et al., 2015) and drinking water treatment plants (Xi et al., 2009; Huang et al., 2011; Chao et al., 2013). Moreover, the ARGs detected in this study encompasses the three major resistance mechanisms, *viz.* exclusion by efflux pumps, antibiotic inactivation and cellular protection mechanisms (Zhu et al. 2013). Hence, detection of these ARGs in both effluent discharged and downstream of the effluent-receiving rivers indicates that WWTP could potentially contribute to the expansion of the resistance reservoir in aquatic environments.

In conclusion, although still at its infancy, metagenomic analysis of total community DNA using direct pyrosequencing without the need for cloning and with long sequence fragments opens up new perspectives in determining the impact of the effluent discharge from the two WWTPs investigated in this study on the antibiotic resistance genes, metabolic potential and bacterial community structures of their receiving rivers. While the sequencing depth in this study was not adequate due to the sequencing platform used, until very recently, 454 pyrosequencing had the advantage of producing the longer average read length with low costs. Rapid development and improvements of next generation sequencing have closed this gap in

the meanwhile on Illumina platforms, increasing both the average read length and the amount of data generated with relatively low costs. Nonetheless, the fact that ARGs, metabolic potential and taxonomic profiles could be detected across all the samples in this study supports the use of whole metagenome shotgun pyrosequencing approach as a relatively simple preliminary investigation for comparing samples exposed to constant fluctuating pressures with unknown microbial community composition and for profiling antibiotic resistance determinants in highly impacted environments. Overall, findings from this study showed the prevalence of a typical WWTP and freshwater bacterial communities. Results from this study indicated a shift in the community structure and metabolic potential of the river ecosystems as a result of the treated effluent discharge from the different WWTP. Additionally, our study suggests, for the first time, the dominance of the elfamycin family of antibiotics in urban WWTPs and their effluent-receiving rivers identifying the possible sources of this family of antibiotics. Findings from this study pave way for a more comprehensive understanding of the microbial ecology of WWTPs and the surrounding aquatic environment and could assist towards effort to safeguard public and environmental health concerns associated with the release of improperly treated wastewater into the surrounding aquatic *milieu*.

4.6 References

Agersø Y, Guardabassi L (2005) Identification of Tet 39, a novel class of tetracycline resistance determinant in *Acinetobacter* spp. of environmental and clinical origin. J Antimicrob Chemother 55:566–569. doi: 10.1093/jac/dki051

Agersø Y, Petersen A (2007) The tetracycline resistance determinant Tet 39 and the sulphonamide resistance gene sulII are common among resistant *Acinetobacter* spp. isolated from integrated fish farms in Thailand. J Antimicrob Chemother 59:23–27. doi: 10.1093/jac/dkl419

Amos GCA, Zhang L, Hawkey PM, et al (2014) Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet Microbiol 171:441–447. doi: 10.1016/j.vetmic.2014.02.017

Andreote FD, Jiménez DJ, Chaves D, et al (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. PLoS One 7:e38600. doi: 10.1371/journal.pone.0038600

Baiser B, Olden JD, Record S, et al (2012) Pattern and process of biotic homogenization in the New Pangaea. Proc Biol Sci 279:4772–4777. doi: 10.1098/rspb.2012.1651

Bäumlisberger M, Youssar L, Schilhabel MB, Jonas D (2015) Influence of a non-hospital medical care facility on antimicrobial resistance in wastewater. PLoS One 10:e0122635. doi: 10.1371/journal.pone.0122635

Berg KA, Lyra C, Sivonen K, et al (2009) High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. ISME J 3:314–325. doi: 10.1038/ismej.2008.110

Brooks BW, Riley TM, Taylor RD (2006) Water quality of effluent-dominated ecosystems: ecotoxicological, hydrological, and management considerations. Hydrobiologia 556:365–379. doi: 10.1007/s10750-004-0189-7

Brown BL, LePrell R V, Franklin RB, et al (2015) Metagenomic analysis of planktonic microbial consortia from a non-tidal urban-impacted segment of James River. Stand Genomic Sci 10:65. doi: 10.1186/s40793-015-0062-5

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. doi: 10.1038/nmeth.3176

Buffie CG, Pamer EG (2013) Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol 13:790–801. doi: 10.1038/nri3535

Chao Y, Ma L, Yang Y, et al (2013) Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci Rep 3:3550. doi: 10.1038/srep03550

Chiaramonte JB, Roberto M do C, Pagioro TA (2014) Large scale distribution of bacterial communities in the upper Paraná River floodplain. Brazilian J Microbiol 45:1187–1197. doi: 10.1590/S1517-83822014000400009

Costa PS, Reis MP, Ávila MP, et al (2015) Metagenome of a microbial community inhabiting a metal-rich tropical stream sediment. PLoS One 10:e0119465. doi: 10.1371/journal.pone.0119465

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485. doi: 10.1186/1471-2105-11-485

Da Silva MLB, Cantão ME, Mezzari MP, et al (2015) Assessment of bacterial and archaeal community structure in Swine wastewater treatment processes. Microb Ecol 70:77–87. doi: 10.1007/s00248-014-0537-8

Department of Water and Sanitation (2015) Green Drop Progress Report. https://www.dwa.gov.za/dir_ws/gds/Docs/DocsDefault.aspx. Accessed 28 Apr 2016

Dijkshoorn L, Nemec A, Seifert H (2007) An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. Nat Rev Microbiol 5:939–951. doi: 10.1038/nrmicro1789

Dinsdale EA, Edwards RA, Hall D, et al (2008) Functional metagenomic profiling of nine biomes. Nature 452:629–632. doi: 10.1038/nature06810

Doxey AC, Kurtz DA, Lynch MDJ, et al (2014) Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. ISME J 9:461–471. doi: 10.1038/ismej.2014.142

Drury B, Rosi-Marshall E, Kelly JJ (2013) Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 79:1897–1905. doi: 10.1128/AEM.03527-12

Du J, Xiao K, Li L, et al (2013) Temporal and spatial diversity of bacterial communities in coastal waters of the South China sea. PLoS One 8:e66968. doi: 10.1371/journal.pone.0066968

Eagar H, Swan G, van Vuuren M (2012) A survey of antimicrobial usage in animals in South Africa with specific reference to food animals. J S Afr Vet Assoc 83:16. doi: 10.4102/jsava.v83i1.16

Economou V, Gousia P (2015) Agriculture and food animals as a source of antimicrobial-resistant bacteria. Infect Drug Resist 8:49–61. doi: 10.2147/IDR.S55778

Ettema TJG, Andersson SGE (2009) The alpha-proteobacteria: the Darwin finches of the bacterial world. Biol Lett 5:429–432. doi: 10.1098/rsbl.2008.0793

Foong CP, Wong Vui Ling CM, González M (2010) Metagenomic analyses of the dominant bacterial community in the Fildes Peninsula, King George Island (South Shetland Islands). Polar Sci 4:263–273. doi: 10.1016/j.polar.2010.05.010

Galm U, Heller S, Shapiro S, et al (2004) Antimicrobial and DNA gyrase-inhibitory activities of novel clorobiocin derivatives produced by mutasynthesis. Antimicrob Agents Chemother 48:1307–1312.

Garnier J, Servais P, Billen G (1992) Bacterioplankton in the Seine River (France): impact of the Parisian urban effluent. Can J Microbiol 38:56–64. doi: 10.1139/m92-009

Gaze WH, Zhang L, Abdouslam NA, et al (2011) Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. ISME J 5:1253–1261. doi: 10.1038/ismej.2011.15

Ghai R, Rodriguez-Valera F, McMahon KD, et al (2011) Metagenomics of the water column in the pristine upper course of the Amazon river. PLoS One 6:e23785. doi: 10.1371/journal.pone.0023785

Gomez-Alvarez V, Teal TK, Schmidt TM (2009) Systematic artifacts in metagenomes from complex microbial communities. ISME J 3:1314–1317. doi: 10.1038/ismej.2009.72

Gücker B, Brauns M, Pusch MT (2006) Effects of wastewater treatment plant discharge on ecosystem structure and function of lowland streams. J North Am Benthol Soc 25:313–329. doi: 10.1899/0887-3593(2006)25[313:EOWTPD]2.0.CO;2

Gullberg E, Cao S, Berg OG, et al (2011) Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog 7:e1002158. doi: 10.1371/journal.ppat.1002158

Guo J, Peng Y, Ni B-J, et al (2015) Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. Microb Cell Fact 14:33. doi: 10.1186/s12934-015-0218-4

Gupta SK, Padmanabhan BR, Diene SM, et al (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. doi: 10.1128/AAC.01310-13

Hall CC, Watkins JD, Georgopapadakou NH (1989) Effects of elfamycins on elongation factor Tu from *Escherichia coli* and *Staphylococcus aureus*. Antimicrob Agents Chemother 33:322–325.

Hamidian M, Holt KE, Pickard D, Hall RM (2016) A small *Acinetobacter* plasmid carrying the *tet39* tetracycline resistance determinant. J Antimicrob Chemother 71:269–271. doi: 10.1093/jac/dkv293

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:1–9.

Howe AC, Jansson JK, Malfatti SA, et al (2014) Tackling soil diversity with the assembly of large, complex metagenomes. Proc Natl Acad Sci 111:4904–4909. doi: 10.1073/pnas.1402564111

Hu M, Wang X, Wen X, Xia Y (2012) Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. Bioresour Technol 117:72–79. doi: 10.1016/j.biortech.2012.04.061

Hu Y, Yang X, Qin J, et al (2013) Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat Commun 4:2151. doi: 10.1038/ncomms3151

Huang J-J, Hu H-Y, Tang F, et al (2011) Inactivation and reactivation of antibiotic-resistant bacteria by chlorination in secondary effluents of a municipal wastewater treatment plant. Water Res 45:2775–2781. doi: 10.1016/j.watres.2011.02.026

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386. doi: 10.1101/gr.5969107

Ilmberger N, Güllert S, Dannenberg J, et al (2014) A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One 9:e106707. doi: 10.1371/journal.pone.0106707

Jeffries TC, Seymour JR, Gilbert JA, et al (2011) Substrate type determines metagenomic profiles from diverse chemical habitats. PLoS One 6:e25173. doi: 10.1371/journal.pone.0025173

Johnning A, Moore ERB, Svensson-Stadler L, et al (2013) Acquired genetic mechanisms of a multiresistant bacterium isolated from a treatment plant receiving wastewater from antibiotic production. Appl Environ Microbiol 79:7256–7263. doi: 10.1128/AEM.02141-13

Jordaan K, Bezuidenhout C (2013) The impact of physico-chemical water quality parameters on bacterial diversity in the Vaal River, South Africa. Water SA 39:385–396. doi: 10.4314/wsa.v39i3.7

Ju F, Guo F, Ye L, et al (2014) Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. Environ Microbiol Rep 6:80–89. doi: 10.1111/1758-2229.12110

Kolmakova O V, Gladyshev MI, Rozanov AS, et al (2014) Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. FEMS Microbiol Ecol 89:442–450. doi: 10.1111/1574-6941.12355

Kristiansson E, Fick J, Janzon A, et al (2011) Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One 6:e17038. doi: 10.1371/journal.pone.0017038

Krzywinski M, Schein J, Birol I, et al (2009) Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645. doi: 10.1101/gr.092759.109

Kümmerer K (2009) Antibiotics in the aquatic environment – A review – Part II. Chemosphere 75:435–441. doi: 10.1016/j.chemosphere.2008.12.006

Li A-D, Li L-G, Zhang T (2015a) Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. Front Microbiol 6:1025. doi: 10.3389/fmicb.2015.01025

Li B, Yang Y, Ma L, et al (2015b) Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J 9:2490–2502. doi: 10.1038/ismej.2015.59

Li D, Yu T, Zhang Y, et al (2010) Antibiotic resistance characteristics of environmental bacteria from an oxytetracycline production wastewater treatment plant and the receiving river. Appl Environ Microbiol 76:3444–3451. doi: 10.1128/AEM.02964-09

Lim YW, Evangelista JS, Schmieder R, et al (2014) Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. J Clin Microbiol 52:425–437. doi: 10.1128/JCM.02204-13

Liu B, Pop M (2009) ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443–D447. doi: 10.1093/nar/gkn656

Liu Z, Huang S, Sun G, et al (2012) Phylogenetic diversity, composition and distribution of bacterioplankton community in the Dongjiang River, China. FEMS Microbiol Ecol 80:30–44. doi: 10.1111/j.1574-6941.2011.01268.x

Luo Y, Mao D, Rysz M, et al (2010) Trends in antibiotic resistance genes occurrence in the Haihe River, China. Environ Sci Technol 44:7220–7225. doi: 10.1021/es100233w

Lupo A, Coyne S, Berendonk TU (2012) Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. Front Microbiol 3:18. doi: 10.3389/fmicb.2012.00018

Manoharan L, Kushwaha SK, Hedlund K, Ahrén D (2015) Captured metagenomics: large-scale targeting of genes based on “sequence capture” reveals functional diversity in soils. DNA Res 22:451–460. doi: 10.1093/dnares/dsv026

Marshall BM, Levy SB (2011) Food animals and antimicrobials: impacts on human health. Clin Microbiol Rev 24:718–733. doi: 10.1128/CMR.00002-11

Marti E, Variatza E, Balcazar JL (2014) The role of aquatic ecosystems as reservoirs of antibiotic resistance. Trends Microbiol 22:36–41. doi: 10.1016/j.tim.2013.11.001

Mason OU, Scott NM, Gonzalez A, et al (2014) Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J 8:1464–1475. doi: 10.1038/ismej.2013.254

McArthur AG, Waglechner N, Nizam F, et al (2013) The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother 57:3348–3357. doi: 10.1128/AAC.00419-13

McKinney M, Lockwood J (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol Evol 14:450–453.

McKinney ML (2008) Effects of urbanization on species richness: A review of plants and animals. Urban Ecosyst 11:161–176. doi: 10.1007/s11252-007-0045-4

Meyer F, Paarmann D, D'Souza M, et al (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386. doi: 10.1186/1471-2105-9-386

Miele A, Goldstein BP, Bandera M, et al (1994) Differential susceptibilities of enterococcal species to elfamycin antibiotics. J Clin Microbiol 32:2016–2018.

Miura Y, Hiraiwa MN, Ito T, et al (2007) Bacterial community structures in MBRs treating municipal wastewater: relationship between community stability and reactor performance. Water Res 41:627–637. doi: 10.1016/j.watres.2006.11.005

Mlejnková H, Sovová K (2010) Impact of pollution and seasonal changes on microbial community structure in surface water. Water Sci Technol 61:2787–2795. doi: 10.2166/wst.2010.080

Munck C, Albertsen M, Telke A, et al (2015) Limited dissemination of the wastewater treatment plant core resistome. Nat Commun 6:8452. doi: 10.1038/ncomms9452

Nemergut DR, Costello EK, Hamady M, et al (2011) Global patterns in the biogeography of bacterial taxa. Environ Microbiol 13:135–144. doi: 10.1111/j.1462-2920.2010.02315.x

Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26:715–721. doi: 10.1093/bioinformatics/btq041

Port JA, Wallace JC, Griffith WC, Faustman EM (2012) Metagenomic profiling of microbial composition and antibiotic resistance determinants in Puget Sound. PLoS One 7:e48000. doi: 10.1371/journal.pone.0048000

Proia L, Osorio V, Soley S, et al (2013) Effects of pesticides and pharmaceuticals on biofilms in a highly impacted river. Environ Pollut 178:220–228. doi: 10.1016/j.envpol.2013.02.022

Qin J, Li R, Raes J, et al (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65. doi: 10.1038/nature08821

R Development Core Team (2007) R: A Language and Environment for Statistical Computing.

Rizzo L, Manaia C, Merlin C, et al (2013) Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ 447:345–360. doi: 10.1016/j.scitotenv.2013.01.032

Roesch LFW, Fulthorpe RR, Riva A, et al (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283–290. doi: 10.1038/ismej.2007.53

Sánchez-Andrea I, Rodríguez N, Amils R, Sanz JL (2011) Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. Appl Environ Microbiol 77:6085–6093. doi: 10.1128/AEM.00654-11

Satinsky BM, Fortunato CS, Doherty M, et al (2015) Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. Microbiome 3:39. doi: 10.1186/s40168-015-0099-0

Sauvain L, Bueche M, Junier T, et al (2013) Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. Aquat Sci 76:33–46. doi: 10.1007/s00027-013-0313-8

Scaria J, Chandramouli U, Verma SK (2005) Antibiotic Resistance Genes Online (ARGO): A database on vancomycin and beta-lactam resistance genes. Bioinformation 1:5–7.

Schmieder R, Edwards R (2012) Insights into antibiotic resistance through metagenomic approaches. Future Microbiol 7:73–89. doi: 10.2217/fmb.11.135

Shanks OC, Newton RJ, Kelty CA, et al (2013) Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl Environ Microbiol 79:2906–2913. doi: 10.1128/AEM.03448-12

Shen SY, Fulthorpe R (2015) Seasonal variation of bacterial endophytes in urban trees. Front Microbiol 6:427. doi: 10.3389/fmicb.2015.00427

Shi P, Jia S, Zhang X-X, et al (2013) Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res 47:111–120. doi: 10.1016/j.watres.2012.09.046

Silva CC, Jesus EC, Torres APR, et al (2010) Investigation of bacterial diversity in membrane bioreactor and conventional activated sludge processes from petroleum refineries using phylogenetic and statistical approaches. J Microbiol Biotechnol 20:447–459.

Smith RJ, Jeffries TC, Roudnew B, et al (2012) Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. Environ Microbiol 14:240–253. doi: 10.1111/j.1462-2920.2011.02614.x

Snaidr J, Amann R, Huber I, et al (1997) Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. Appl Environ Microbiol 63:2884–2896.

Storteboom H, Arabi M, Davis JG, et al (2010) Identification of antibiotic-resistance-gene molecular signatures suitable as tracers of pristine river, urban, and agricultural sources. Environ Sci Technol 44:1947–1953. doi: 10.1021/es902893f

Tang K, Liu K, Jiao N, et al (2013) Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. PLoS One 8:e72958. doi: 10.1371/journal.pone.0072958

Uroz S, Ioannidis P, Lengelle J, et al (2013) Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation. PLoS One 8:e55929. doi: 10.1371/journal.pone.0055929

Vogeley L, Palm GJ, Mesters JR, Hilgenfeld R (2001) Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. J Biol Chem 276:17149–17155. doi: 10.1074/jbc.M100017200

Wakelin SA, Colloff MJ, Kookana RS (2008) Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. Appl Environ Microbiol 74:2659–2668. doi: 10.1128/AEM.02348-07

Wang Z, Zhang X-X, Huang K, et al (2013) Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. PLoS One 8:e76079. doi: 10.1371/journal.pone.0076079

WHO (2014) Antimicrobial resistance: global report on surveillance 2014. 257.

Wilke A, Bischof J, Gerlach W, et al (2015) The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Res 44:D590–D594. doi: 10.1093/nar/gkv1322

Wolf H, Chinali G, Parmeggiani A (1974) Kirromycin, an inhibitor of protein biosynthesis that acts on elongation factor Tu. Proc Natl Acad Sci U S A 71:4910–4914.

Wu H, Zhao H, Wen C, et al (2012) A comparative study of bacterial community structures in the sediments from brominated flame retardants contaminated river and non-contaminated reservoir. African J Microbiol Res 6:3248–3260. doi: 10.5897/AJMR11.1640

Xi C, Zhang Y, Marrs CF, et al (2009) Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Appl Environ Microbiol 75:5714–5718. doi: 10.1128/AEM.00382-09

Xia S, Duan L, Song Y, et al (2010) Bacterial community structure in geographically distributed biological wastewater treatment reactors. Environ Sci Technol 44:7391–7396. doi: 10.1021/es101554m

Yang J (2014) Viral pathogens in clinical samples by use of a metagenomic approach. In: Nelson KE (ed) Encycl. Metagenomics, 1st edn. Springer, New York, pp 1–6

Yu K, Zhang T (2012) Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7:e38183. doi: 10.1371/journal.pone.0038183

Yu K, Zhang T (2013) Construction of customized sub-databases from NCBI-nr database for rapid annotation of huge metagenomic datasets using a combined BLAST and MEGAN approach. PLoS One 8:e59831. doi: 10.1371/journal.pone.0059831

Zhang T, Zhang X-X, Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One 6:e26041. doi: 10.1371/journal.pone.0026041

Zhang Y, Cong J, Lu H, et al (2015) Soil bacterial diversity patterns and drivers along an elevational gradient on Shennongjia Mountain, China. Microb Biotechnol 8:739–746. doi: 10.1111/1751-7915.12288

Zheng W, Zhang Z, Liu C, et al (2015) Metagenomic sequencing reveals altered metabolic pathways in the oral microbiota of sailors during a long sea voyage. Sci Rep 5:9131. doi: 10.1038/srep09131

Zhou CE, Smith J, Lam M, et al (2007) MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:D391–D394. doi: 10.1093/nar/gkl791

4.7 Figures

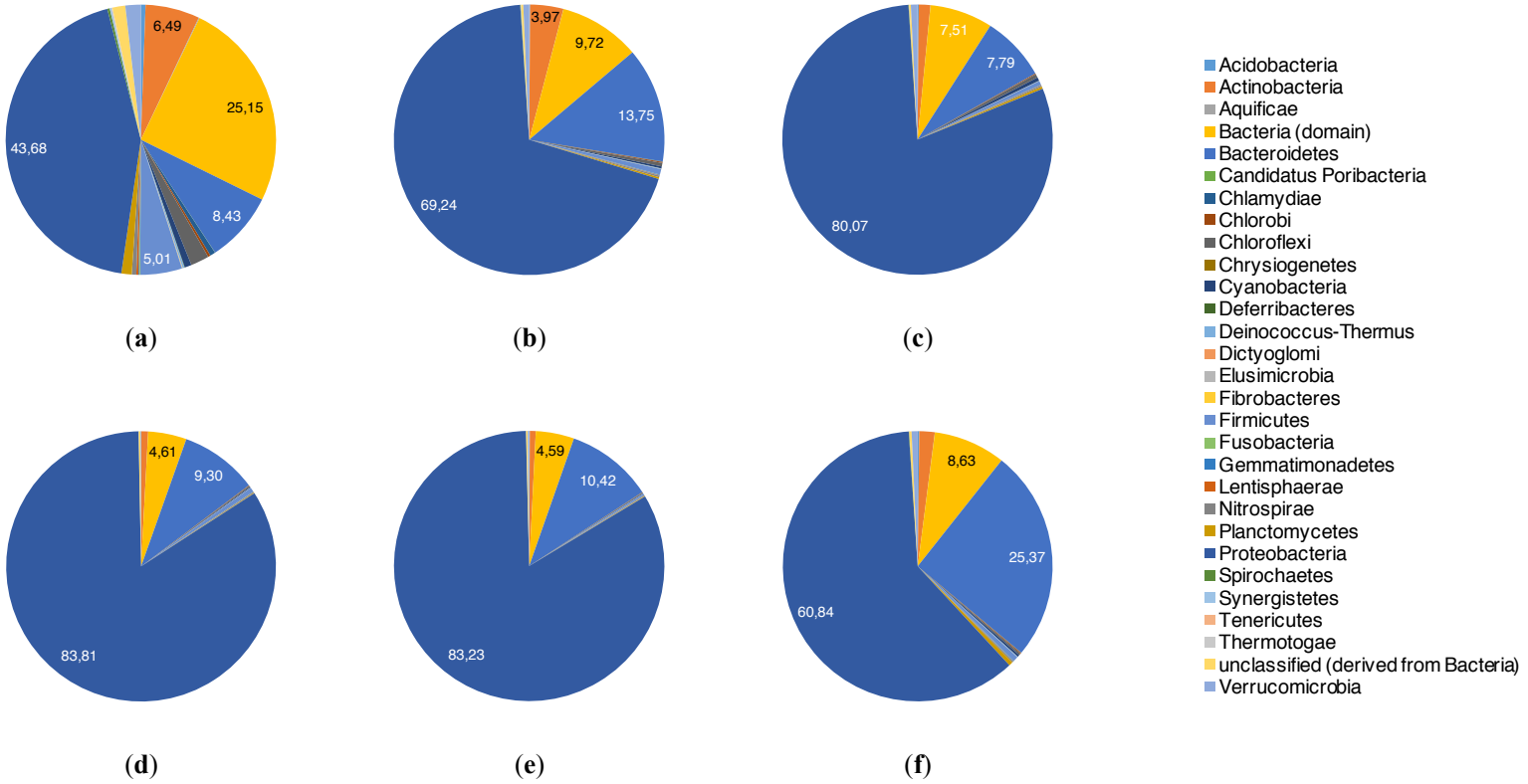


Figure 4.1: Relative abundances of dominant phyla in the metagenome of (**a**) Plant A effluent discharged, (**b**) downstream of river receiving effluent from Plant A, (**c**) upstream of river receiving effluent from Plant A, (**d**) Plant B effluent discharged, (**e**) downstream of river receiving effluent from Plant B and (**f**) upstream of of river receiving effluent from Plant B as annotated by MG-RAST. Relative abundance represents the number of reads affiliated with that phyla divided by the total reads assigned for the bacterial domain. Several phyla dominating the domain in the sample is indicated on the chart as percentages.

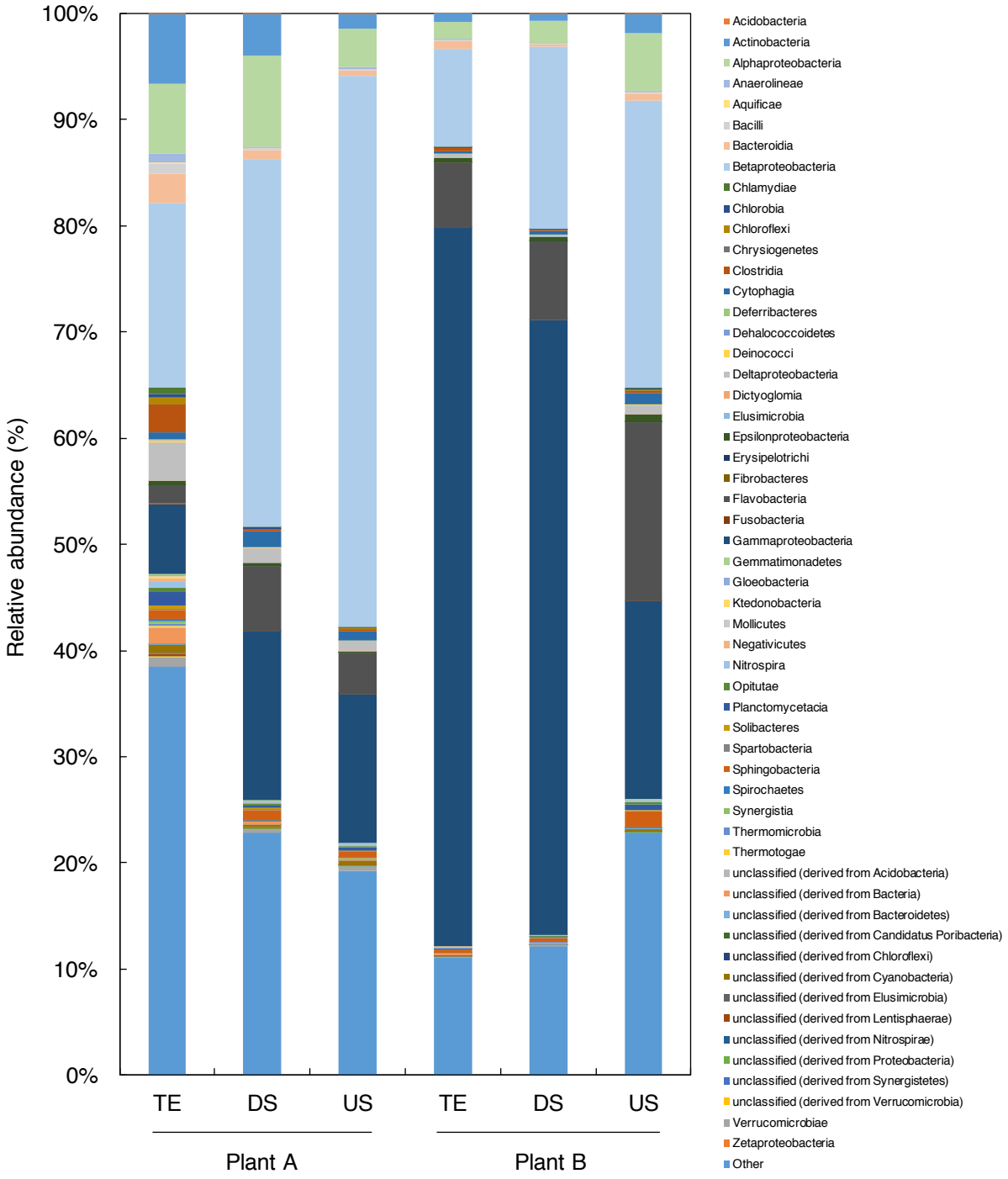


Figure 4.2: Relative abundances of dominant classes observed in the metagenome of the treated effluents (TE) of both WWTPs, and upstream (US) and downstream (DS) of their receiving rivers as annotated by MG-RAST. Relative abundance represents the number of reads affiliated with that class divided by the total reads assigned for the bacterial domain. The "Other" category in the figure represents assignments to the class level rather than potential novel class (represented as "unclassified derived from").

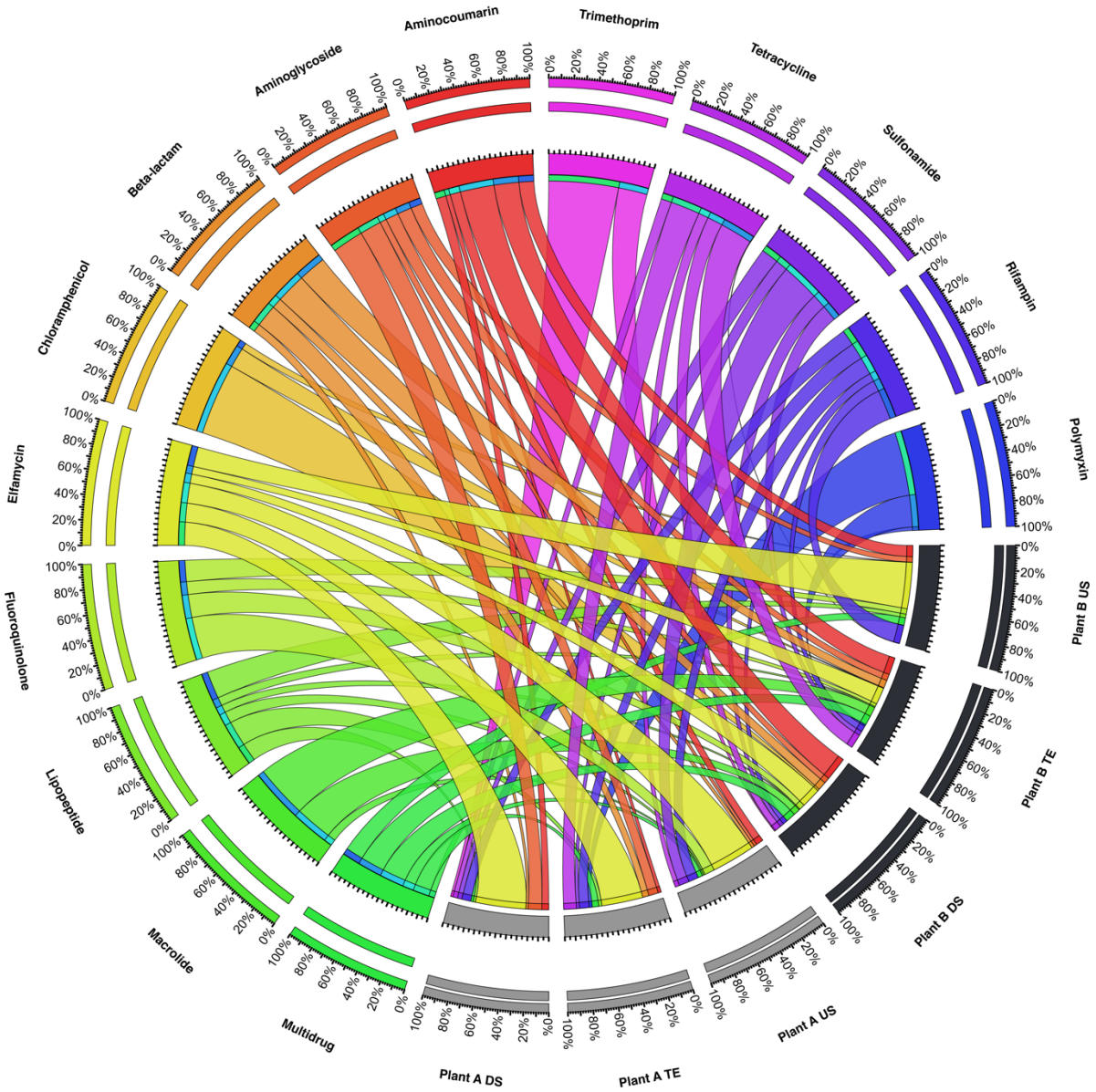


Figure 4.3: Distribution of the relative abundance of antibiotic resistance genes (ARGs) types in the metagenome of the treated effluents (TE) of both WWTPs, and upstream (US) and downstream (DS) of their receiving rivers as visualized *via* Circos. Relative abundance values were calculated by dividing the number of annotated ARGs by the total number of ARGs detected in the metagenomes. Each antibiotic resistance type is represented by a specific ribbon colour, and the width of the outer ring for each ribbon represents the percentage relative abundance of ARGs in the associated metagenomes. Plant A and Plant B associated metagenomes has been coloured in grey and black, respectively.

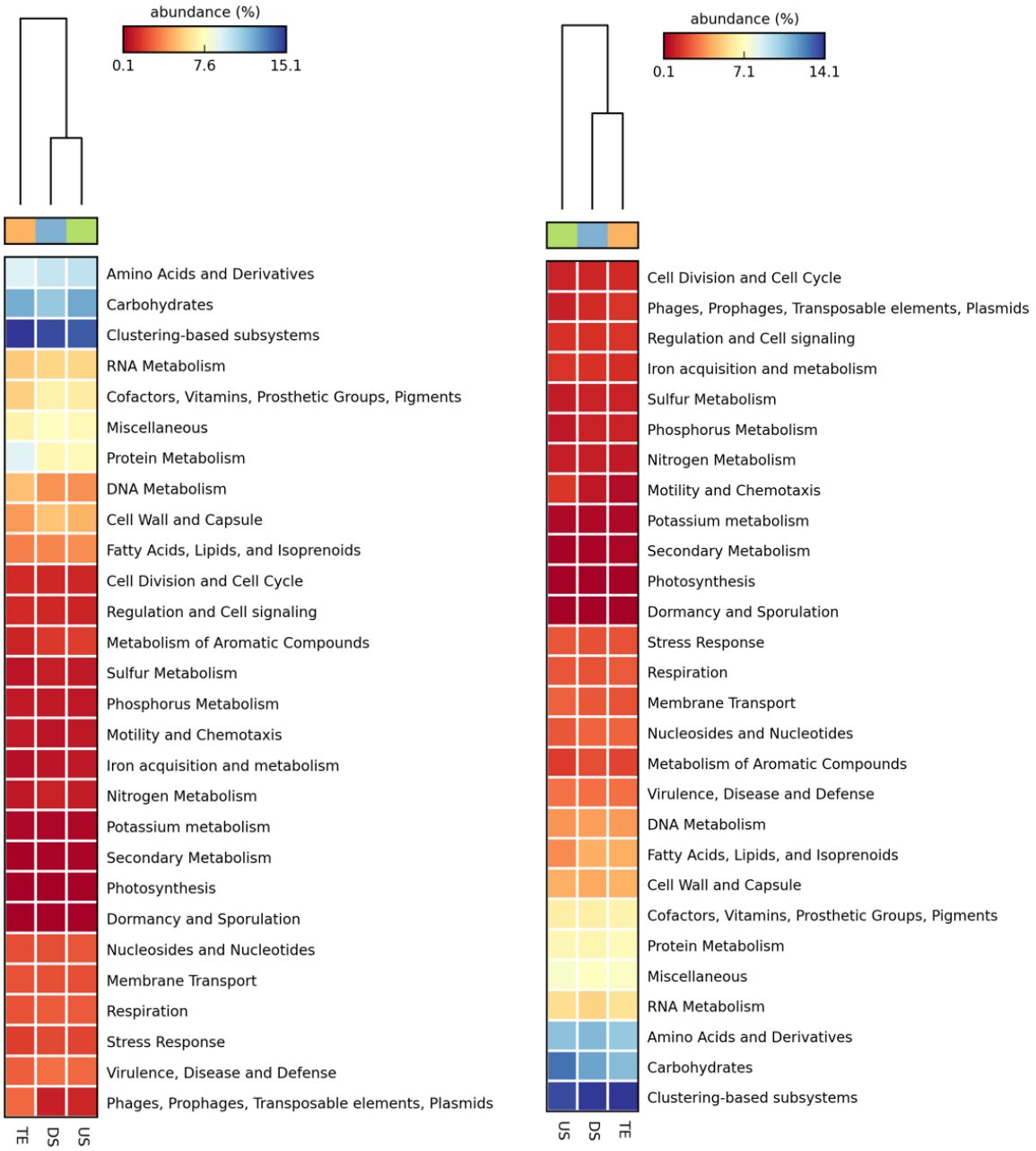
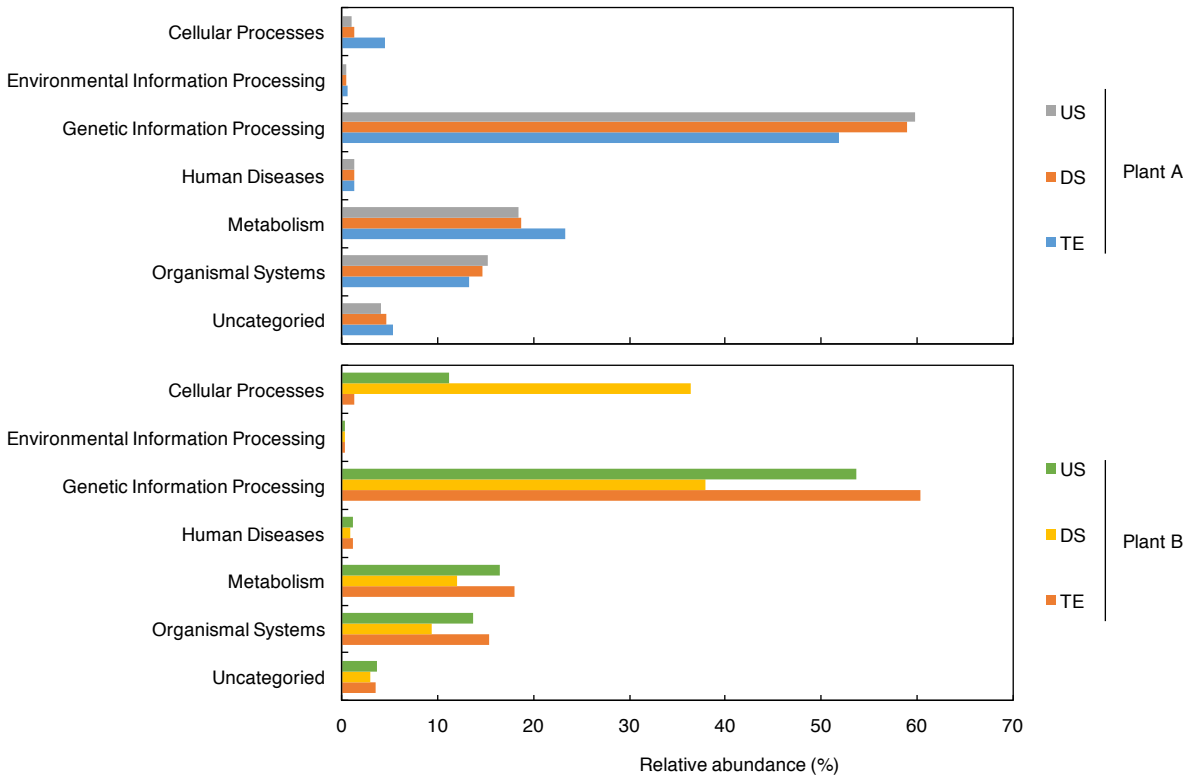
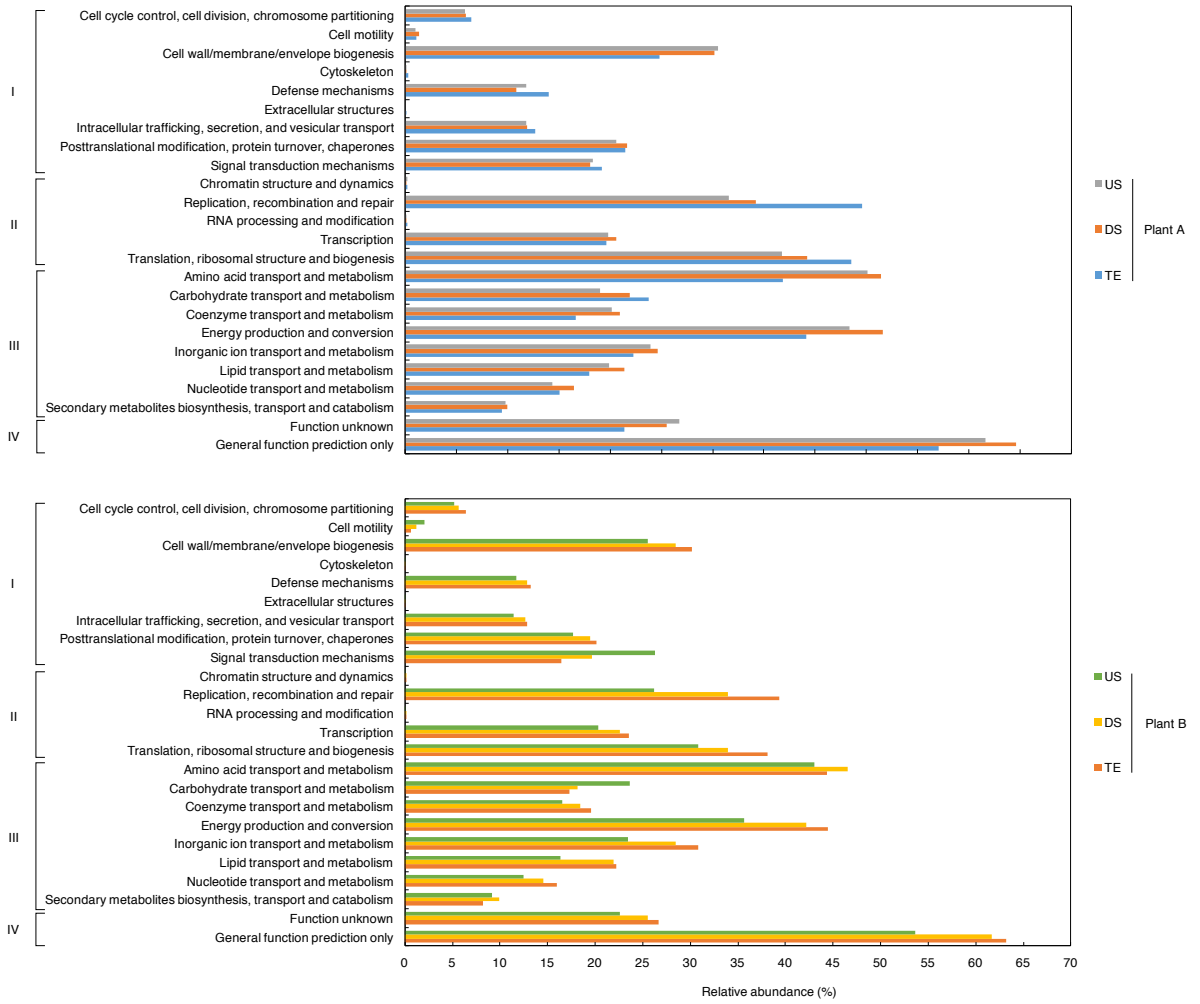


Figure 4.4: Heatmap depicting the distribution and relative abundance of predicted proteins assigned to 28 functional level 1 SEED subsystems in the metagenome of (**a**) Plant A treated effluent (TE), downstream (DS) and upstream (US) of receiving river of Plant A TE and (**b**) Plant B TE, DS and US of receiving river of Plant B TE as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a subsystem divided by the total reads assigned to all SEED subsystems.



(**a**)

Figure 4.5: Distribution and relative abundance of predicted proteins assigned to (**a**) 7 major KEGG hierarchies (**b**) 24 COG families in the metagenomes of the treated effluent (TE) of both WWTPs, downstream (DS) and upstream (US) of the respective receiving rivers as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a classification/family divided by the total reads assigned to all KEGG/COG. COG families are divided into 4 major groups; I, Cellular processes and signalling; II, Information storage and processing; III, Metabolism; IV, Poorly characterized.



(b)

Figure 4.5 (cont.): Distribution and relative abundance of predicted proteins assigned to (**a**) 7 major KEGG hierarchies (**b**) 24 COG families in the metagenomes of the treated effluent (TE) of both WWTPs, downstream (DS) and upstream (US) of the respective receiving rivers as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a classification/family divided by the total reads assigned to all KEGG/COG. COG families are divided into 4 major groups; I, Cellular processes and signalling; II, Information storage and processing; III, Metabolism; IV, Poorly characterized.

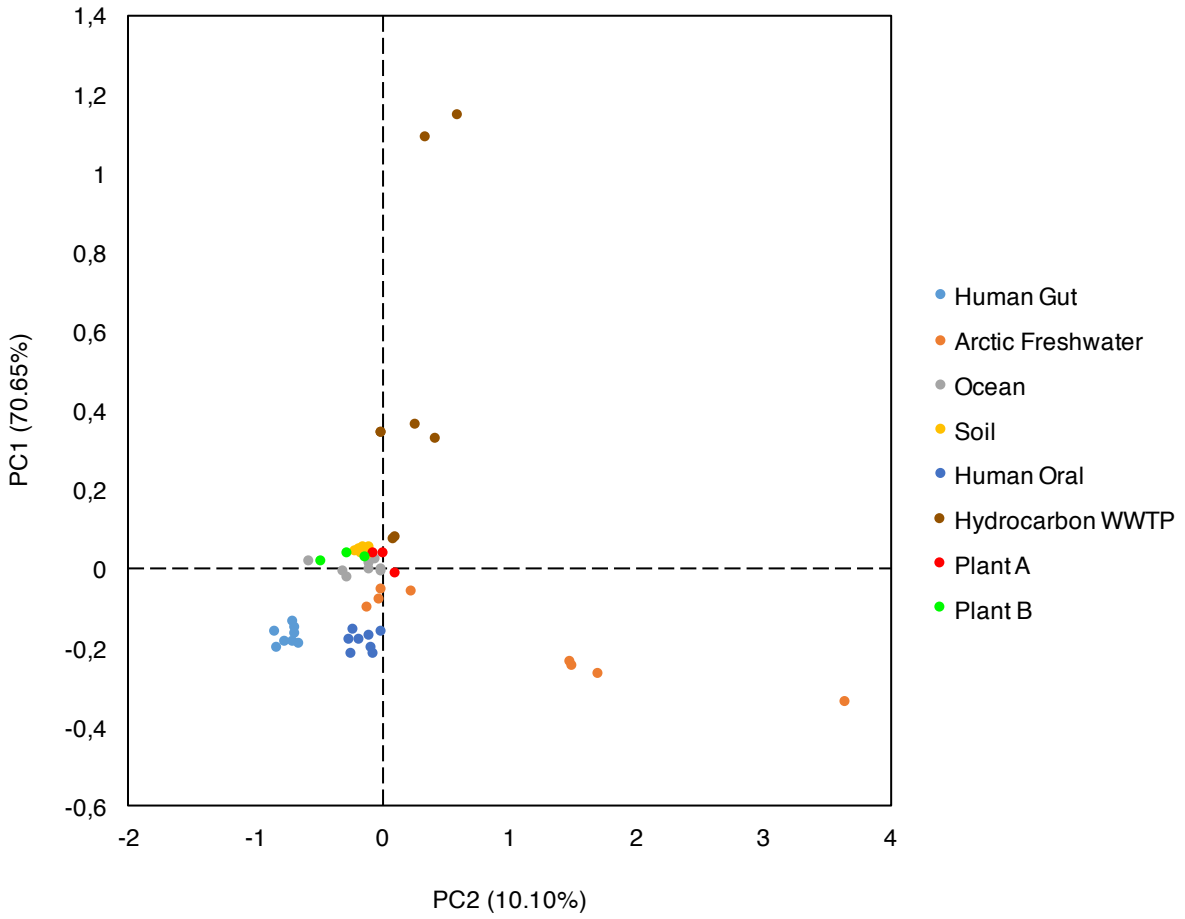


Figure 4.6: Principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity coefficients comparing the metagenomes of the investigated wastewater treatment plants with 5 other ecosystems (human oral, human gut, ocean, soil and freshwater) with normalized amount of annotated reads in the SEED subsystems. Metagenomes of the different ecosystems were analysed in MG-RAST to avoid variations in analysis using the publicly available data on MG-RAST (their information is provided in Table S4.2).

4.8 Supplementary materials

Table S4.1: Characterization of the 454 pyrosequenced libraries generated from the treated effluents of wastewater treatment plants (WWTPs) and the respective receiving river water samples analysed in this study.

Characteristics	Plant A			Plant B		
	WWTP Effluent	**River Downstream**	**River Upstream**	**WWTP Effluent**	**River Downstream**	**River Upstream**
# raw reads	227,021	115,483	169,225	180,557	148,377	233,173
Total size (bp)	143,321,109	73,823,069	96,909,864	110,996,053	93,930,993	136,123,317
Average reads (bp)	631	639	573	615	633	584
# reads uploaded to MG-RAST	227,021	115,483	169,225	180,557	148,377	233,173
MG-RAST QC						
# reads before QC	227,021	115,483	169,225	180,557	148,377	233,173
Total size before QC (bp)	143,321,109	73,823,069	96,909,864	110,996,053	93,930,993	136,123,317
Average length before QC (bp)	631	639	573	615	633	584
# reads removed during dereplication (bp)	11,511	5,388	10,567	2,031	7,808	13,986
# after QC	194,388	99,617	147,605	147,749	127,426	201,260
Total size after QC (bp)	41,085,936	20,168,242	37,429,256	27,636,500	26,061,955	49,921,741
Average length after QC (bp)	211	203	254	187	205	248
MG-RAST Annotations						
# reads identified as RNA	32,299	15,356	20,112	23,065	19,395	29,977
# reads predicted ORFs	175,950	90,658	142,899	131,523	118,258	192,916
# reads assigned to SEED	63,773	67,359	136,518	103,755	102,227	155,558
# reads assigned to KEGG	19,038	19,829	40,538	31,009	30,783	45,177
# reads assigned to COG	32,332	33,676	67,537	50,763	52,556	85,143

Table S4.2: Basic information summary of 48 publicly available metagenomic datasets from MG-RAST in six ecosystems (human gut, human oral, ocean, soil, freshwater and wastewater treatment plants) used in this study.

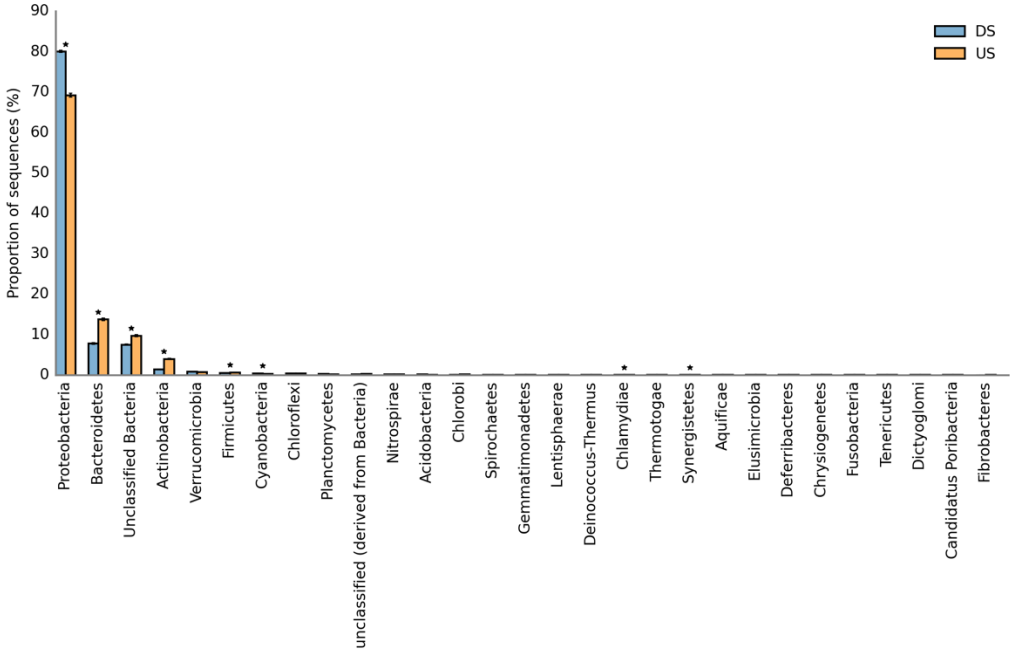
Label/Biomes	MG-RAST ID	Metagenome size (bp)	# Reads	Habitat Type	Location	Sequencing Method
Human Oral_1	4447943,3	142,374,233	339,503	human-associated	Valencia, Spain	454
Human Oral_2	4447192,3	77,538,485	204,218	human-associated	Valencia, Spain	454
Human Oral_3	4447103,3	203,711,161	464,594	human-associated	Valencia, Spain	454
Human Oral_4	4447102,3	100,125,112	244,881	human-associated	Valencia, Spain	454
Human Oral_5	4447101,3	129,851,692	295,072	human-associated	Valencia, Spain	454
Human Oral_6	4447971,3	37,519,874	97,722	human-associated	Valencia, Spain	454
Human Oral_7	4447970,3	27,669,924	70,503	human-associated	Valencia, Spain	454
Human Oral_8	4447903,3	123,266,763	306,74	human-associated	Valencia, Spain	454
Human Gut_1	4440452,7	54,632,274	229,857	human-associated	St, Louis, United States	454
Human Gut_2	4440616,3	174,824,393	507,928	human-associated	St, Louis, United States	454
Human Gut_3	4440611,3	103,097,122	526,727	human-associated	St, Louis, United States	454
Human Gut_4	4440826,3	124,768,172	499,499	human-associated	St, Louis, United States	454
Human Gut_5	4440824,3	100,520,072	414,497	human-associated	St, Louis, United States	454
Human Gut_6	4440639,3	93,430,618	440,521	human-associated	St, Louis, United States	454
Human Gut_7	4440461,5	105,923,024	522,134	human-associated	St, Louis, United States	454
Human Gut_8	4440613,3	102,979,597	312,665	human-associated	St, Louis, United States	454
Ocean_1	4443702,3	47,289,202	209,073	aquatic	Bergen, Norway	454
Ocean_2	4443707,3	31,359,337	135,033	aquatic	Bergen, Norway	454
Ocean_3	4443708,3	8,571,342	38,22	aquatic	Bergen, Norway	454
Ocean_4	4443703,3	30,991,689	134,915	aquatic	Bergen, Norway	454
Ocean_5	4443709,3	26,982,195	116,192	aquatic	Bergen, Norway	454
Ocean_6	4443705,3	68,187,679	304,02	aquatic	Bergen, Norway	454
Ocean_7	4443704,3	59,316,369	344,216	aquatic	Bergen, Norway	454
Ocean_8	4443706,3	38,021,523	162,871	aquatic	Bergen, Norway	454
Soil_1	4445996,3	116,821,792	312,444	soil	Navada, United States	454
Soil_2	4445993,3	133,555,260	352,417	soil	Navada, United States	454

Table S4.2 (cont.): Basic information summary of 48 publicly available metagenomic datasets from MG-RAST in six ecosystems (human gut, human oral, ocean, soil, freshwater and wastewater treatment plants) used in this study.

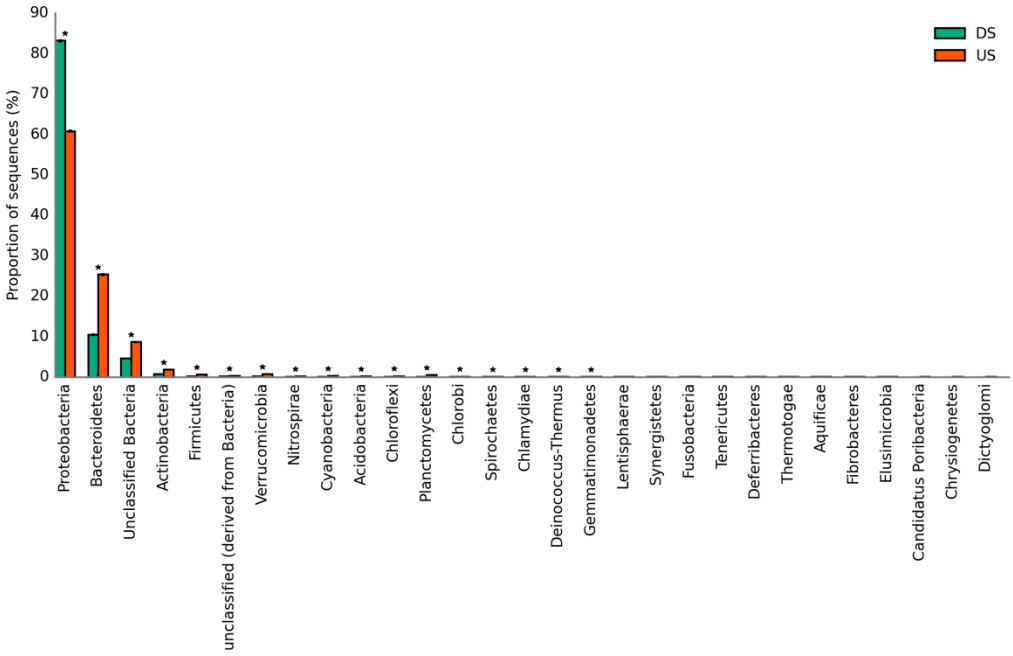
Label/Biomes	MG-RAST ID	Metagenome size (bp)	# Reads	Habitat Type	Location	Sequencing Method
Soil_3	4445994,3	254,548,462	683,082	soil	Navada, United States	454
Soil_4	4445990,3	219,117,356	583,724	soil	Navada, United States	454
Soil_5	4450750,3	87,160,647	239,933	soil	Navada, United States	454
Soil_6	4450752,3	76,860,743	233,279	soil	Navada, United States	454
Soil_7	4451103,3	397,257,248	1,040,697	soil	Navada, United States	454
Soil_8	4451104,3	347,578,191	998,484	soil	Navada, United States	454
Arctic Freshwater_1	4443683,3	101,310,476	100,085	freshwater	Antarctica	sanger
Arctic Freshwater_2	4443680,3	9,622,231	9,672	freshwater	Antarctica	sanger
Arctic Freshwater_3	4443679,3	9,755,315	10,042	freshwater	Antarctica	sanger
Arctic Freshwater_4	4443681,3	54,929,769	54,446	freshwater	Antarctica	sanger
Arctic Freshwater_5	4443682,3	284,069,722	283,663	freshwater	Antarctica	sanger
Arctic Freshwater_6	4443685,3	28,413,296	28,481	freshwater	Antarctica	sanger
Arctic Freshwater_7	4443686,3	101,573,008	103,058	freshwater	Antarctica	sanger
Arctic Freshwater_8	4443687,3	95,664,001	95,521	freshwater	Antarctica	sanger
Hydrocarbon WWTP_1	4507688,3	116,596,410	186,74	terrestrial	Inniskillen, Canada	454
Hydrocarbon WWTP_2	4507689,3	155,694,182	241,209	terrestrial	Inniskillen, Canada	454
Hydrocarbon WWTP_3	4509513,3	29,379,862	88,526	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_4	4509514,3	41,742,688	109,108	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_5	4509515,3	50,103,178	124,879	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_6	4509516,3	38,880,264	111,957	terrestrial	San Juan, United States	454
Hydrocarbon WWTP_7	4523101,3	423,985,802	548,46	terrestrial	Fort McMurray, Canada	454
Hydrocarbon WWTP_8	4523103,3	340,253,804	436,518	terrestrial	Fort McMurray, Canada	454

Table S4.3: Relative abundance of the microbial populations in the treated effluent discharge from two full-scale wastewater treatment plants, and upstream and downstream of their effluent-receiving river as annotated and assigned by the MG-RAST pipeline. Relative abundance is reported as percentages and represents the number of reads affiliated with that domain divided by the total reads assigned for all domain.

Domain	Plant A			Plant B		
	WWTP Effluent	River Downstream	River Upstream	WWTP Effluent	River Downstream	River Upstream
Bacteria	93.32	98.37	98.94	99.44	99.52	98.69
Eukaryota	4.37	1.16	0.80	0.24	0.29	1.02
Archaea	1.47	0.26	0.15	0.17	0.11	0.12
Viruses	0.47	0.11	0.07	0.11	0.04	0.09
Unclassified sequences	0.37	0.09	0.05	0.03	0.03	0.07
Other sequences	<0.01	0.01	<0.01	0.01	<0.01	<0.01



(**a**)



(**b**)

Figure S4.1: Comparative taxonomic profile of downstream (DS) and upstream (US) samples of the river receiving treated effluent discharge by (**a**) Plant A and (**b**) Plant B at phylum taxonomic level. Annotation was determined by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Asterisks indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP.

(**a**)

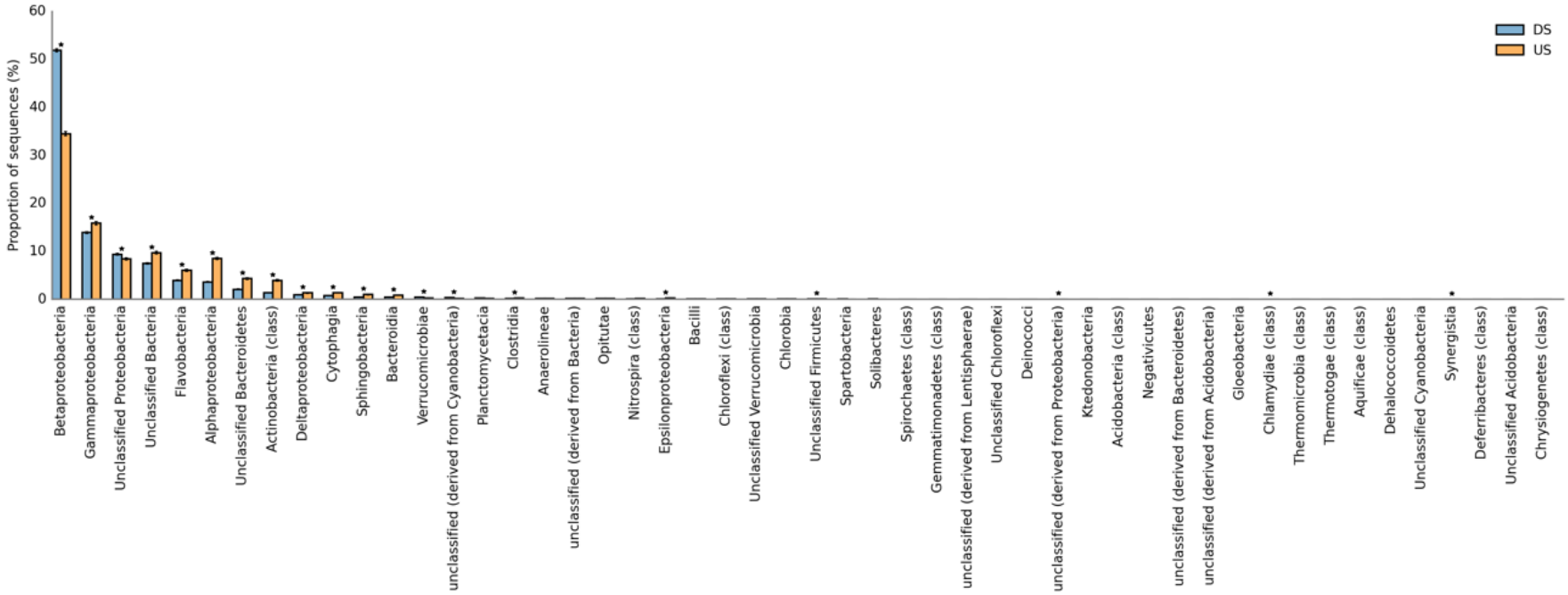


Figure S4.2: Comparative taxonomic profile of downstream (DS) and upstream (US) samples of the river receiving treated effluent discharge by (**a**) Plant A and (**b**) Plant B at class taxonomic level. Annotation was determined by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Asterisks indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP.

(**b**)

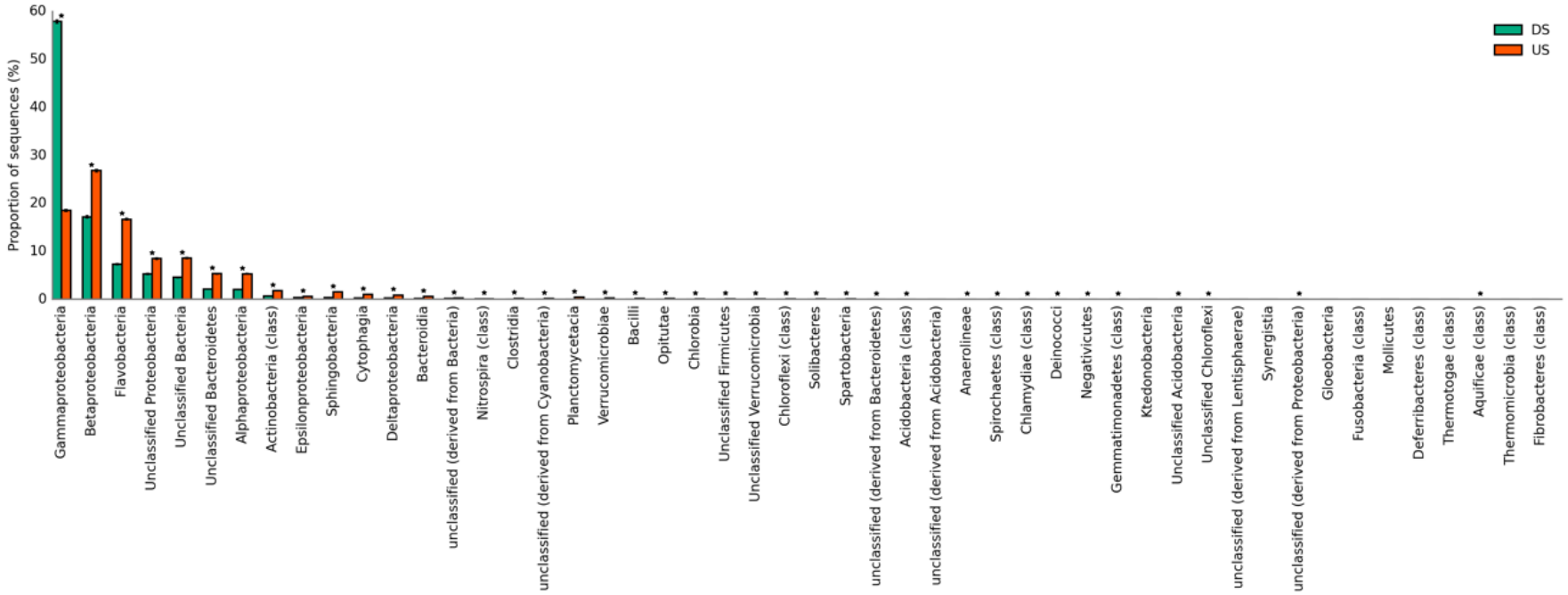


Figure S4.2: Comparative taxonomic profile of downstream (DS) and upstream (US) samples of the river receiving treated effluent discharge by (**a**) Plant A and (**b**) Plant B at class taxonomic level. Annotation was determined by the MG-RAST pipeline and assigned with the lowest common ancestor algorithm. Asterisks indicate phyla with significant differences (q<0.05) in abundance between the wastewater samples determined in STAMP.

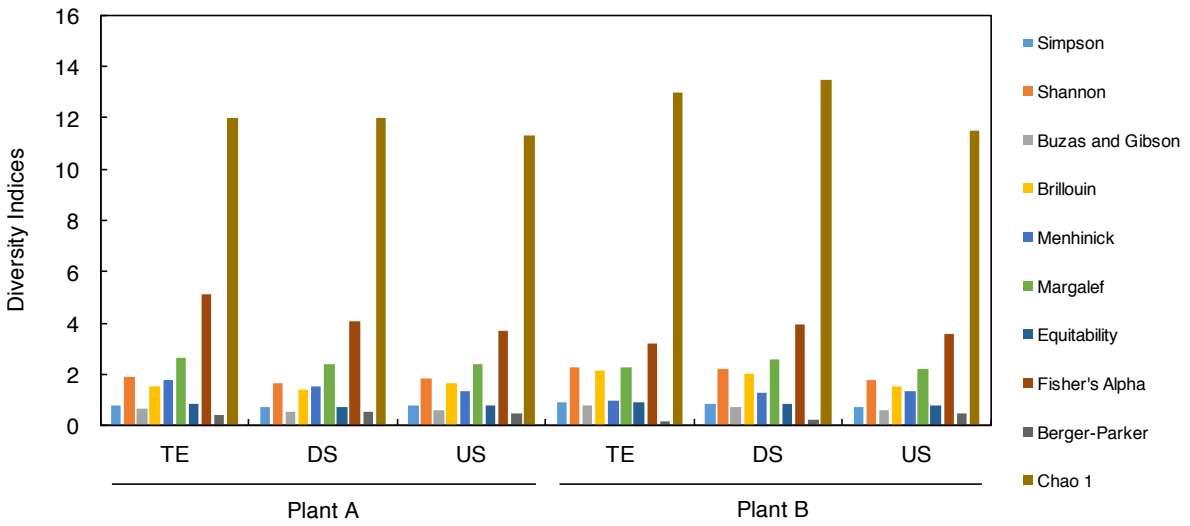


Figure S4.3: Diversity indices of the ARGs types calculated for treated effluent discharged (TE) from both WWTPs, and downstream (DS) and upstream (US) of the corresponding effluent-receiving rivers.

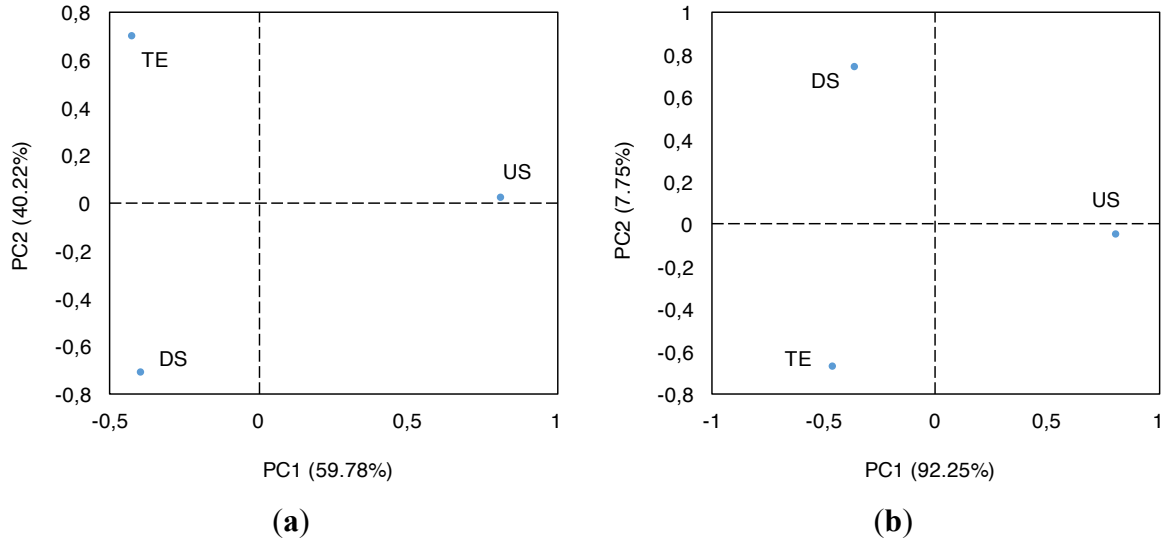


Figure S4.4: Principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity coefficients of different ARGs classes determined for treated effluent discharged (TE) downstream (DS) and upstream (US) of the effluent-receiving rivers for (**a**) Plant A and (**b**) Plant B.

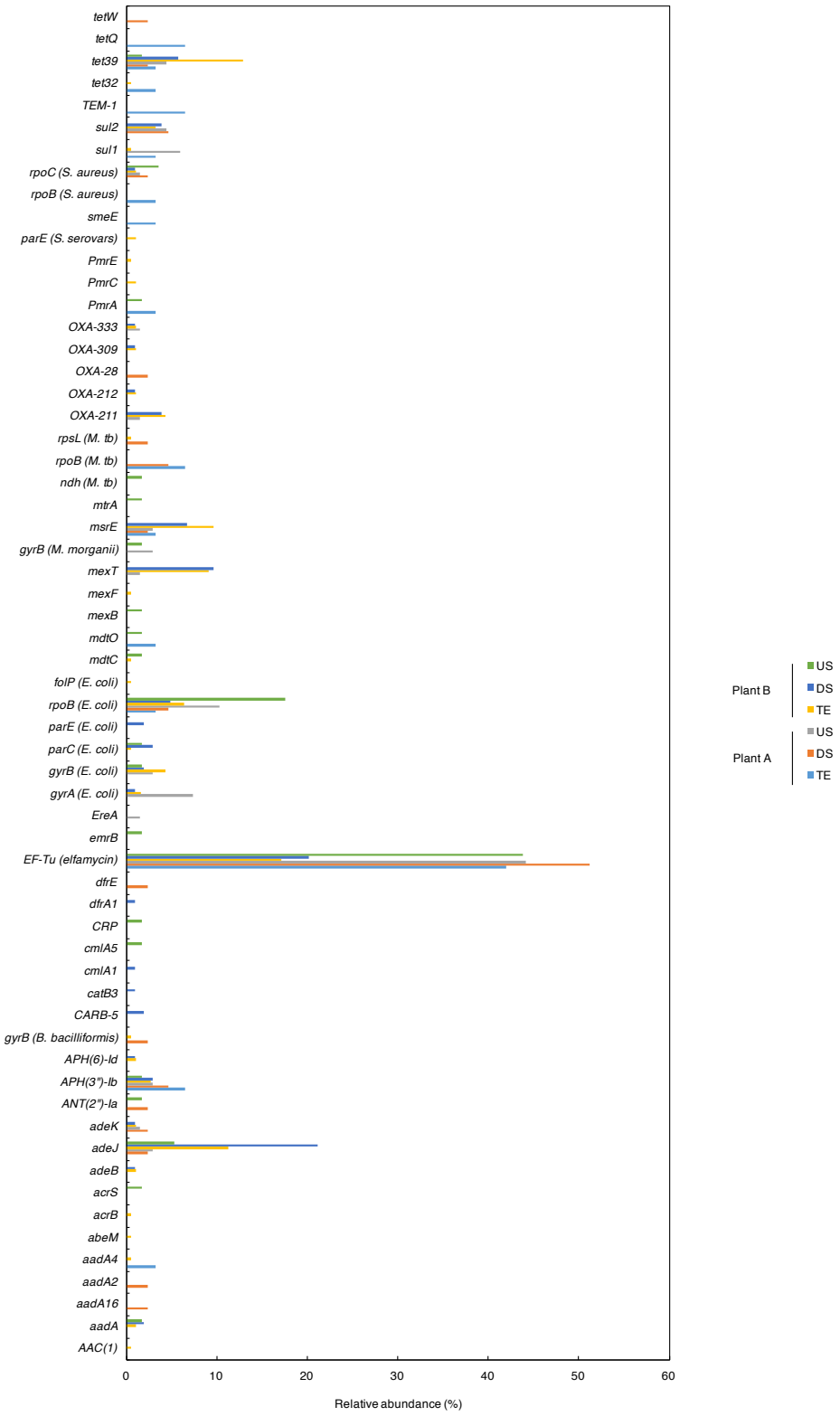


Figure S4.5: Comparison of the diversity of antibiotic resistance genes (ARGs) in Plant A and Plant B across all metagenomic samples. Relative abundance values were calculated by dividing the number of annotated as ARGs to the total number of ARGs detected in the metagenome. ARGs were annotated by BLASTx against the CARD protein database. (*Staphylococcus aureus*, *S. aureus*; *Salmonella enterica*, *S. serovars*; *Mycobacterium tuberculosis*, *M. tb*; *Escherichia coli*, *E. coli; Bartonella bacilliformis, B. bacilliformis*).

Table S4.4: Distribution and relative abundance of predicted proteins assigned to level 1 SEED subsystems metagenomes of treated effluent discharged (TE) of Plant A and in downstream (DS) and upstream (US) of the receiving river as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a subsystem divided by the total reads assigned to all SEED subsystems.

SEED Level 1 Subsystems	TE	DS	US
Amino Acids and Derivatives	7,25	8,10	8,13
*Carbohydrates	8,54	8,91	8,01
Cell Division and Cell Cycle	1,73	1,62	1,64
*Cell Wall and Capsule	3,77	4,32	4,70
Clustering-based subsystems	16,28	15,28	15,56
Cofactors, Vitamins, Prosthetic Groups, Pigments	5,85	7,05	7,12
DNA Metabolism	5,57	4,30	4,22
Dormancy and Sporulation	0,20	0,22	0,20
Fatty Acids, Lipids, and Isoprenoids	2,79	2,94	2,78
Iron acquisition and metabolism	0,50	0,73	0,67
Membrane Transport	3,33	3,13	3,14
Metabolism of Aromatic Compounds	1,34	1,91	1,76
Miscellaneous	7,72	8,22	8,54
Motility and Chemotaxis	0,92	0,84	0,79
*Nitrogen Metabolism	1,26	1,41	1,61
Nucleosides and Nucleotides	2,85	2,99	2,71
*Phages, Prophages, Transposable elements, Plasmids	3,21	1,53	1,43
Phosphorus Metabolism	0,74	0,74	0,75
Photosynthesis	0,11	0,19	0,13
Potassium metabolism	0,37	0,39	0,40
Protein Metabolism	9,82	8,62	8,51
Regulation and Cell signalling	1,56	1,42	1,52
Respiration	3,16	3,39	3,47
RNA Metabolism	4,63	4,73	4,71
Secondary Metabolism	0,32	0,42	0,40
Stress Response	2,25	2,31	2,47
Sulphur Metabolism	0,94	1,14	1,34
Virulence, Disease and Defence	3,00	3,16	3,29

*indicate statistically significant differences (q<0.05) between DS and US assignments

Table S4.5: Distribution and relative abundance of predicted proteins assigned to level 1 SEED subsystems in metagenomes of treated effluent discharged (TE) of Plant B and in downstream (DS) and upstream (US) of the effluent-receiving river as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a subsystem divided by the total reads assigned to all SEED subsystems.

SEED Level 1 Subsystems	TE	DS	US
*Amino Acids and Derivatives	8,83	9,06	8,55
*Carbohydrates	7,83	8,52	9,80
*Cell Division and Cell Cycle	1,75	1,61	1,41
Cell Wall and Capsule	4,09	4,01	4,06
*Clustering-based subsystems	14,97	14,93	14,78
Cofactors, Vitamins, Prosthetic Groups, Pigments	6,85	6,71	6,42
*DNA Metabolism	4,30	4,26	4,10
Dormancy and Sporulation	0,17	0,21	0,22
*Fatty Acids, Lipids, and Isoprenoids	3,11	3,24	2,66
Iron acquisition and metabolism	1,26	1,31	1,53
*Membrane Transport	3,14	3,23	3,55
*Metabolism of Aromatic Compounds	1,76	2,00	1,75
*Miscellaneous	8,22	8,20	8,25
*Motility and Chemotaxis	0,57	0,91	1,64
Nitrogen Metabolism	1,25	1,41	1,39
*Nucleosides and Nucleotides	3,14	3,11	2,75
*Phages, Prophages, Transposable elements, Plasmids	2,03	1,72	1,32
*Phosphorus Metabolism	0,92	0,87	0,71
Photosynthesis	0,13	0,12	0,13
Potassium metabolism	0,46	0,44	0,44
Protein Metabolism	8,20	7,85	7,87
Regulation and Cell signalling	1,75	1,60	1,72
Respiration	3,06	2,93	2,99
*RNA Metabolism	4,76	4,39	4,58
*Secondary Metabolism	0,48	0,41	0,32
Stress Response	2,51	2,54	2,61
*Sulphur Metabolism	1,30	1,28	1,15
Virulence, Disease and Defence	3,14	3,16	3,29

*indicate statistically significant differences (q<0.05) between DS and US assignments

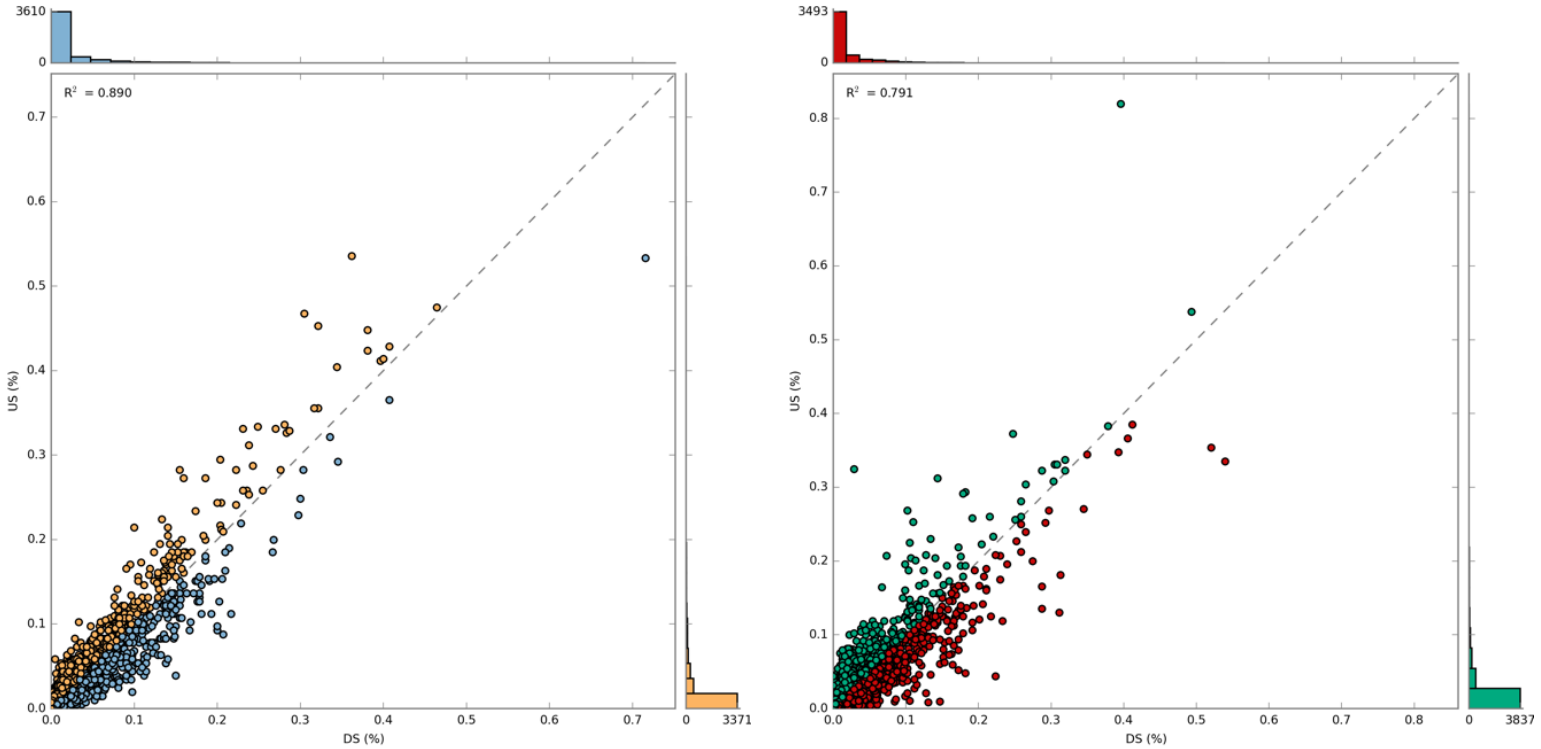
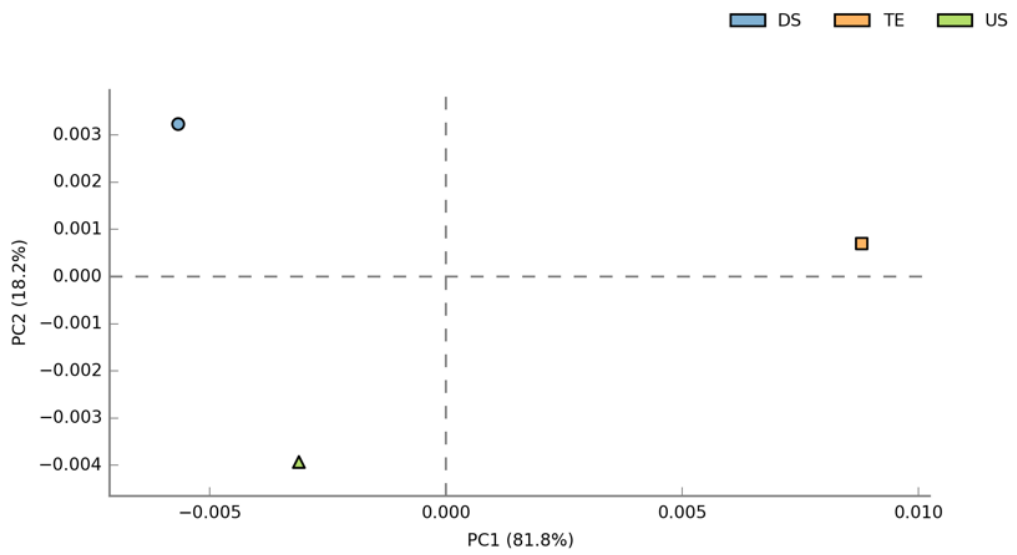
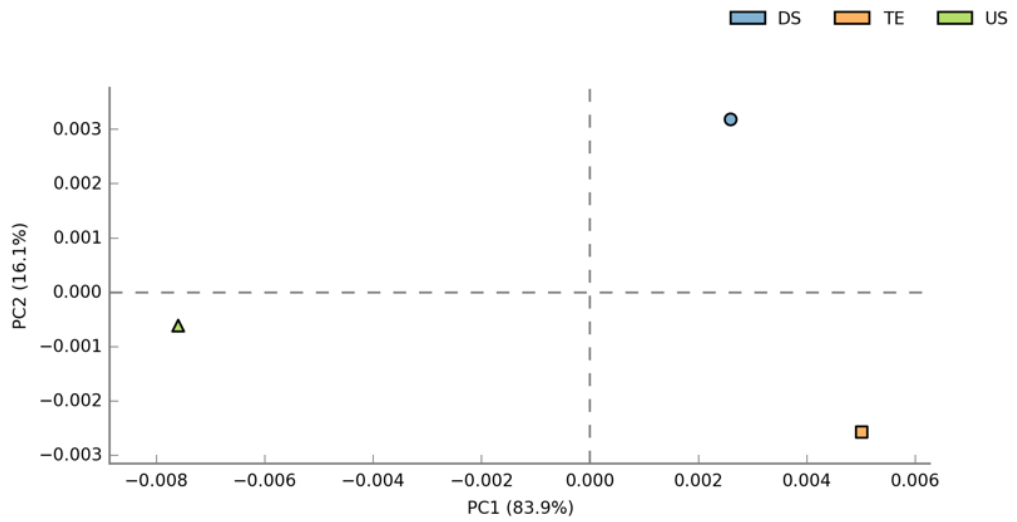


Figure S4.6: Comparison between the SEED subsystem assignment of predicted proteins at the functional level of complexity in (**a**) downstream (DS) and upstream (US) of river receiving effluent discharged by Plant A, and (**b**) DS and US samples of river receiving treated effluent discharged by Plant B as determined using the STAMP software package.

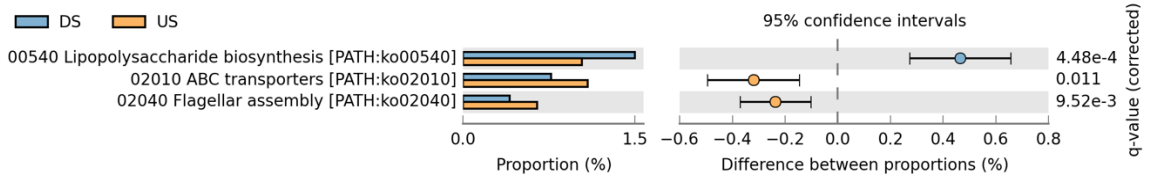


(**a**)



(**b**)

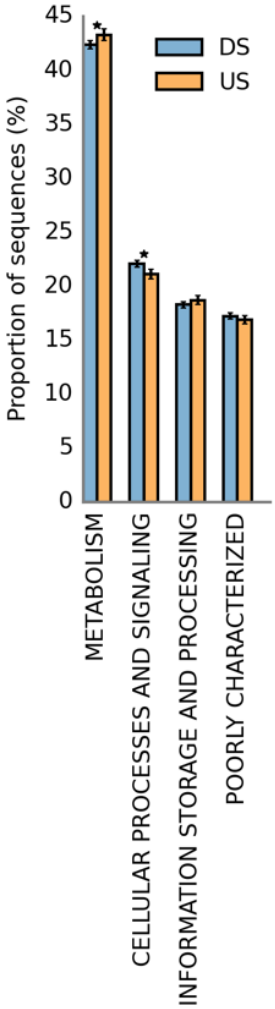
Figure S4.7: Principal component analysis (PCA) of the SEED level 1 subsystems assigned in MG-RAST for treated effluent discharged (TE) and upstream (US) and downstream (DS) of the effluent-receiving rivers for (**a**) Plant A and (**b**) Plant B metagenomes.



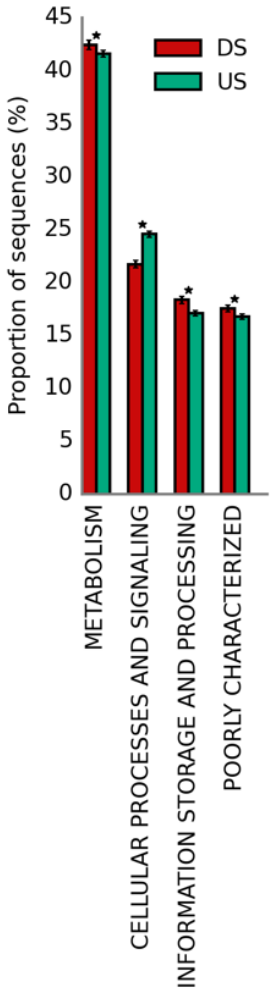
(**a**)

(**b**)

Figure S4.8: Chart depicting post-hoc confidence interval plots (>95%) based on Fisher's exact test parameters including SEED level 3 carried out in STAMP and assigned in MG-RAST for treated effluent discharged (TE) and upstream (US) and downstream (DS) of the effluent-receiving rivers for (**a**) Plant A and (**b**) Plant B.



(**a**)



(**b**)

Figure S4.9: Comparison between the relative abundance of predicted proteins assigned to COG families determined using the STAMP software package for treated effluent discharged (TE) and upstream (US) and downstream (DS) of the effluent-receiving rivers for (**a**) Plant A and (**b**) Plant B. Asterisks indicate significant differences (q<0.05) in abundance between the river samples.

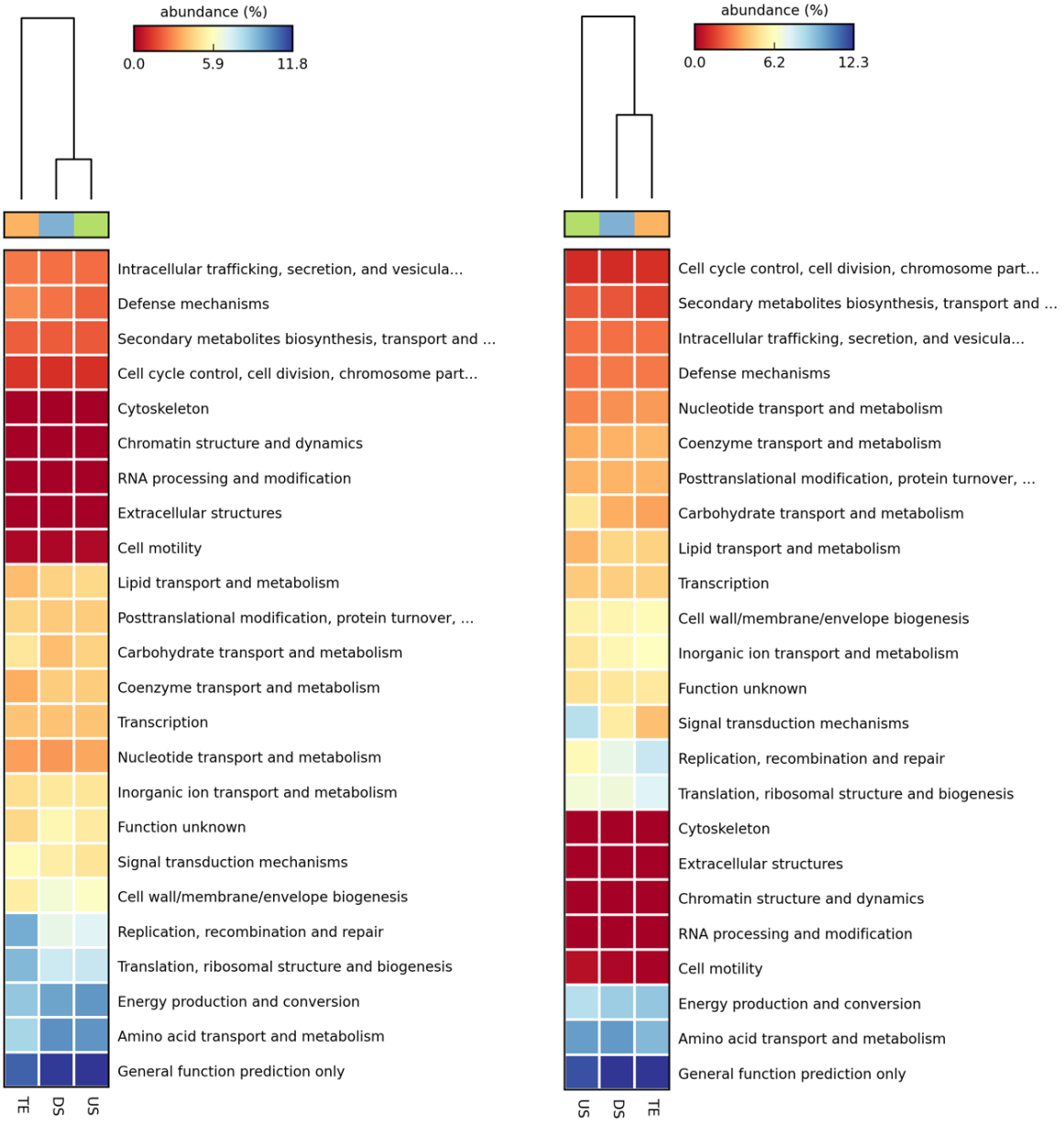


Figure S4.10: Heatmap depicting the distribution and relative abundance of predicted proteins assigned to the level 2 complexity COG families in the metagenome of (**a**) Plant A treated effluent discharged (TE), downstream (DS) and upstream (US) of effluent-receiving river receiving river of Plant A and (**b**) Plant B TE, DS and US of effluent-receiving river of Plant B as annotated by MG-RAST. Relative abundance is reported as percentages and represents the number of reads affiliated with a family divided by the total reads assigned to all COG families.

Chapter Five

Manuscript formatted and submitted to *World Journal of Microbiology and Biotechnology*

5 Bacterial community dynamics of two rivers receiving treated urban wastewater effluents: a T-RFLP assessment

5.1 Abstract

Effluents from wastewater treatment plants (WWTPs) represent a significant proportion of the receiving river water body. In this study, we investigated the effects of treated wastewater effluent discharge from two full-scale WWTPs on the microbial diversity of the receiving river bodies. Total bacterial DNA was extracted from the effluent and river water samples and analyzed by the terminal-restriction fragment length polymorphism (T-RFLP) fingerprinting of the 16S rRNA gene. Different bacterial fingerprint patterns observed for all sample sites were compared. The overall effect of the WWTP effluent on the urban river was the observed indistinguishable communities from upstream to downstream of the river. However, a significant decrease (p<0.01) in bacterial community diversity and composition was observed downstream of the river, relative to the upstream. Conversely, for the suburban river, no significant decrease in bacterial diversity and composition was observed downstream of the river subsequent to receiving treated wastewater effluent. Results from this study suggest that depending on the wastewater effluent quality and receiving water body, treated wastewater effluent has the potential to reduce the natural bacterial diversity in river ecosystems which may have serious implications on the higher order ecosystems.

Keywords: Bacterial community diversity; Environment; River; T-RFLP; Wastewater effluent

5.2 Introduction

Urban wastewater treatment plants (WWTPs) are one of the most common systems used for treating domestic, industrial and hospital wastewater. In highly urbanized areas with high population densities, WWTPs can be large and numerous. For example, the KwaZulu-Natal province of South Africa has 11 districts and is serviced by 140 WWTPs with a total design capacity of 1090.8 Ml/day (Department of Water and Sanitation 2015). The treatment of wastewater rely heavily on the activities of complex microbial communities for effective reduction of various physicochemical parameters and pathogen load before disposal into the receiving water bodies or in some cases re-use in agriculture and aquaculture sectors (World Health Organization 2006). Majority of WWTPs frequently discharge the final treated effluent into nearby rivers, and in many cases wastewater effluent discharge make up a significant proportion of the receiving water body (Brooks et al. 2006; Gücker et al. 2006). However, rivers in urban and suburban areas are often heavily impacted by the poorly treated wastewater discharge from WWTPs, and result in severe reduction in the water quality leading to detrimental effects on the health of these aquatic ecosystems (Carey and Migliaccio 2009). Furthermore, it is impossible for the characteristics of the final treated effluent to match the characteristics of the water in the receiving bodies despite the level of treatment and compliance to water quality standards set at local and government levels (Drury et al. 2013). Thus, effluent from WWTPs has the potential to significantly alter the physicochemical and microbial properties of the receiving ecosystem.

Assessments of river health is largely based on the use of indicator bacteria and other microorganisms to provide a more illuminating condition of the aquatic ecosystem due to their ubiquitous presence, high abundance and continued exposure to the water columns as well as

their major involvement in biogeochemical transformations (Lawrence et al. 2005). Furthermore, the impacts of stress and disturbance on the bacterial and other microbial communities can lead to the understanding of the implications of these stress factors on ecosystem functions and processes as well as biodiversity in higher order community structures. However, only a minor fraction of bacteria have been isolated in pure cultures on appropriate media *in vitro*, accounting for <1% of the natural diversity in the environment (Vartoukian et al. 2010). Consequently, studies have utilized cultivation-independent genetic fingerprinting techniques such as denaturing gradient gel electrophoresis (DGGE), terminal-restriction fragment length polymorphism (T-RFLP), single-strand conformation polymorphism, random amplified polymorphic DNA, amplified ribosomal DNA restriction analysis, length heterogeneity PCR, and ribosomal intergenic spacer analysis, to provide insights into the diversity and structure of microbial communities (Ahmad et al. 2011). For example, 16S rRDA-DGGE was used to assess the effect of wastewater effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow in South Australia (Wakelin et al. 2008). They found that significant shifts in bacterial community structures were associated with alteration of the sediment's physicochemical properties. Furthermore, amplified ribosomal DNA restriction analysis was used to demonstrate the reduction of the bacterial and archaeal diversity in an anaerobic sequencing-batch biofilm reactor for the treatment of industrial wastewater (Sarti et al. 2012).

In this study, T-RFLP fingerprinting technique was used to investigate the changes in bacterial community diversity and structure at selected treatment stages of two different full-scale WWTPs as well as the effects of the final treated effluent discharge on the receiving river bodies in the city of Durban, South Africa. Plant A's receiving river is located in a less densely

populated suburban area, whilst Plant B's receiving river is located in a highly urbanized area, with both rivers receiving effluent from a separate large WWTP. Results from this study may help to extend our knowledge of the changes in the complex bacterial community diversity and composition during wastewater treatment processes as well as the effect of the final treated effluent on the diversity and composition of bacterial community in receiving river bodies.

5.3 Materials and Methods

5.3.1 Study description and sample collections

In this study, two full-scale WWTPs, hereon designated as Plant A and Plant B, which discharges their final treated effluent directly into nearby rivers in the city of Durban, South Africa, were chosen for investigation. Plant A has a capacity of 70 megaliters/day with an operational capacity of 96% and uses the activated sludge and diffused air liquid technologies with gravity thickening, anaerobic digestion and belt press dewatering sludge technologies. The plant discharges the final effluent into a nearby river in a suburban location. Plant B has a capacity of 25 megaliters/day with an operational capacity of 76% and use the activated sludge liquid technology with anaerobic digestion and belt press dewatering sludge technologies. The plant discharges the final effluent into a nearby river in heavily urbanized location (Department of Water and Sanitation 2015). Furthermore, Plant A receives a mixture of domestic and industrial wastewater, whilst Plant B receives a mixture of domestic, industrial and hospital wastewaters. Three pre-designated sampling points within the plants *viz.*, raw influent received by the plants (RI), the treated effluent before disinfection by chlorination (BD), and the final effluent discharged (FE) were chosen to allow for comparison of bacterial communities during

the treatment process. To determine the shifts in bacterial community diversity as a result of the final effluent discharged into the receiving river, two pre-designated points along the rivers *viz.*, downstream (DS) and upstream (US), approximately 500 m in opposite directions from the final effluent discharge point from the WWTPs were selected. All samples were collected in 5 L plastic bottles pre-sterilized with 70% (v/v) ethanol and rinsed with 4 L of the sample at the various sampling sites prior to collection. Upon collection, the samples were transported on ice back to the laboratories within 3 h and stored at 4°C prior to DNA extraction which took place within 24 h.

5.3.2. Total DNA extraction, PCR amplification and restriction digestion

Total DNA of the bacterial populations were extracted from the samples using PowerWater™ DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. This kit is specifically designed for isolating bacterial DNA from environmental water samples and includes inhibitor removal technology aimed at removing humic acid and other organic matter commonly found in environmental samples that can interfere with downstream analyses. The quantity and purity of the isolated DNA was determined using the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, USA).

Amplification of the universally conserved regions of 16S ribosomal DNA of the bacterial genomes (approx. 1.5 kb) was carried out by PCR using primer set F27 (5′-AGA GTT TGA TCM TGG CTC AG-3′) and R1492 (5′-TAC GGY TAC CTT GTT ACG ACT-3′), labelled with 6-carboxyfluorescein (FAM) and hexachlorofluorescein (HEX) at the 5′ end, respectively (Disayathanoowat et al. 2012). PCR consisted of 1 ng of template DNA, 0.4 mmol of each

primer, 2.5 µl of 10× reaction buffer (Southern Cross Biotechnology, South Africa), 1 µl of 25 mM $MgCl_2$ (Southern Cross Biotechnology, South Africa), 0.5 µl of 10 mM dNTP Mix (Thermo Fisher Scientific, USA), 2 U of 5U/µl SuperTherm *Taq* (Southern Cross Biotechnology, South Africa) and nuclease-free water (Thermo Fisher Scientific, USA) to make up a 25 µl reaction mixture. PCR amplification was done in triplicate per DNA sample in the T100 Thermo Cycler (Bio-Rad, USA) under the following conditions: initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 45 s, annealing at 55°C for 1 min, extension 72°C for 1 min, with a final extension at 72°C for 10 min. PCR products (3 µl) were mixed with 0.6 µl of 6× loading dye (Thermo Fisher Scientific, USA) and then loaded onto a 1% (w/v) Sekem LE Agarose gel (Lonza, USA) with a 1 kb Plus DNA Ladder (Thermo Fisher Scientific, USA). Subsequently, the gel was stained with 0.2 mg/ml ethidium bromide (Bio-Rad, USA) and visualized under UV light in G:BOX F3 system (Cambridge, United Kingdom) to determine whether amplification of correct product sizes had occurred. Purification of remaining PCR products was done using GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA) with an elution product of 30 µl. Quality and concentrations of purified PCR products was determined with the $OD_{260/280}$ values of 2 µl of PCR products and ranged from 24.3 to 102.1 ng DNA/µl.

Purified PCR products was simultaneously digested using three FastDigest restriction enzymes, *HhaI*, *MspI* and *RsaI*, at 1 U each, 2 µl of 10× FastDigest Buffer, 10 µl of purified PCR products and nuclease-free water to make up a 30 µl reaction mixture. Reaction digestion mixtures were incubated at 37°C for 4 h to ensure complete digestion. Purification of restriction digestion products was done using GeneJET PCR Purification Kit with an elution product of 30 µl. All enzymes, reagents and kits were purchased from Thermo Fisher Scientific, USA.

5.3.3 Capillary electrophoresis and sizing of T-RFLP profiles

To determine the length of terminal-restriction fragments (T-RFs), all 30 µl of purified digested products was mixed with the LIZ-labelled GeneScan™-500 internal size standards (Life Technologies, USA). The samples were denatured at 95°C for 3 min and then placed on ice for 5 min. T-RFLP capillary electrophoresis was carried out on an automated sequencer, the ABI Genetic Analyser 3500xL (Applied Biosystems, UK), by Inqaba Biotechnical Industries (Pretoria, South Africa) and T-RFs generated by the sequencer were analysed using the size-calling software GeneMapper™ v.4.0 (Applied Biosystems, UK). AFLP option was selected in the software with 10 to 1000 bp set as the analysis range and quantification was carried out in the advanced mode using global southern method algorithm. A T-RF in a T-RFLP profile was called by the software if the minimum peak height was above the noise observed (10, 50 and 60 for blue, green and yellow relative fluorescence units (RFU), respectively). Sizing tables containing raw peak positions, peak height and peak area per profile were exported from GeneMapper™ for further analysis.

5.3.4 Consensus T-RFLP profiles and community structure analysis

Raw datasets was transformed into appropriate format required by the T-RFLP excel tool (Fredriksson et al. 2014a). For quality control of the different replicate profiles, normalization, alignment, alignment correction, systematic differences detection was initially determined based on methodology described by Kaplan et al. (2001) implemented in the tool. Subsequently, normalization, alignment, alignment correction, and systematic differences detection of generated consensus profiles for each sample point was computed. The final consensus profile for each sample site generated was used for further analysis. In this analysis, the relative peak

area from the consensus profiles data matrix was processed. Relative peak area was calculated by dividing the individual peak area of each T-RF in a sample by the sum of all T-RF peak areas in the sample (Wang et al. 2010). This was done in order to compensate for differences in the quantity of PCR product and subsequent T-RFs profile intensity variations among samples.

In order to graphically represent the structure of a bacterial community in each T-RFLP profile, Pareto-Lorenz (PL) evenness curves were constructed as previously described (Marzorati et al. 2008; Wang et al. 2010). For the various T-RFLP profiles, each T-RF in the profile was ranked from high to low based on their relative area abundance. The cumulative proportions of relative area abundance were plotted against the respective cumulative proportions of each T-RF peak, to generate a convex curve. The more the PL curve deviates from the theoretical line representing perfect evenness (i.e., the 45° diagonal), the less evenness is observed in the community structure of the profile. In this study, the PL curves were also evaluated according to the Pareto's principle where the cumulative y-axis value corresponding to the 20% level on the x-axis was evaluated (Possemiers et al. 2004).

5.3.5 Diversity indices and nonmetric multidimensional scaling analysis

Raw datasets was transformed into appropriate format required by the T-RFLP excel tool (Fredriksson et al. 2014a). For quality control of the different replicate profiles, normalization, alignment, alignment correction and systematic differences detection was determined as described previously. The triplicate profiles for each sample site generated were used for further analysis. In this analysis, the relative peak area from the data matrix was processed as

described previously. All univariate and multivariate statistical analyses was performed with the 'vegan' package (Shen and Fulthorpe 2015) from the program R (R Development Core Team 2007).

Richness (*S*) was determined by the presence or absence of unique T-RFs in each of the profiles. Shannon's diversity index (*H'*) was determined using the equation 1, Pielou's Evenness index (*J'*) was calculated using equation 2, Simpson's diversity index (*D*) was determined using equation 3 and Simpson's reciprocal diversity index (*1/D*) was calculated using equation 4 as described by Tipayno et al. (2012).

$$H' = -\sum(p_i)(\log p_i) \quad (1)$$

$$J' = \mathrm{H'}/\log(\mathrm{S}) \quad (2)$$

$$D = 1 - \sum \mathrm{p_i}^2 \quad (3)$$

$$1/D = \frac{1}{\sum \mathrm{p_i}^2} \quad (4)$$

In these formulas, *S* was richness, p_i was the relative area abundance of the T-RFs to the specific profile and *log* was the natural logarithm with base 10. Both Shannon and Simpson's indices were used because they respond differently to changes in rare and abundant species. The Shannon diversity index has increased sensitivity to changes in rare species whereas Simpson's reciprocal diversity index is more sensitive to changes in abundant species (Blackwood et al. 2007). Comparison of diversity indices between the sample site were done by one-way analysis of variance (ANOVA) using IBM SPSS v.22 (Armonk, New York: IBM Corp.).

Pair-wise comparisons of the profiles for each sample sites was done using the Bray-Curtis distance coefficient (Legendre and Legendre 1998) and visualised using nonmetric

multidimensional scaling (NMDS) using the 'metaMDS' method in the 'vegan' package. A one-way analysis of similarity (ANOSIM) was then used to examine the statistical significance of differences observed among the T-RFs at the different sample sites. Permutation for the analysis was set at 1000.

5.4 Results

5.4.1 Richness, community structure and evenness

In this study, a total of 10 samples from two different WWTPs and receiving river bodies were analysed. Richness, as derived from the number of unique T-RF peaks detected from each of the consensus profiles that were within the range of the size markers used (35 to 500 bp) is shown in Table 5.1 for the various sample points. The overall highest richness was observed in Plant A with a total of 91 T-RFs, whilst the lowest was observed in Plant B with a total of 43 T-RFs. In Plant A, highest richness was observed in the FE of the WWTP, followed by DS and US of the suburban river, then RI and BD of the WWTP. Conversely, in Plant B, the highest richness was observed US of the urban river, followed by RI and BD of the WWTP, and DS of the urban river, then FE of the WWTP.

The relative abundance of each unique T-RF of the bacterial community from the T-RFLP profiles showed that the bacterial community was not stable across all the samples collected (Figure 5.1). For Plant A, relative abundances of 36-, 80-, 100-, 118-, 119-, 120-, 201-, 204-, 429- and 486-bp T-RFs varied greatly at the different treatment stages and from river surface

waters. The 120-bp T-RF was present at a relative abundance of 3.87% from the raw influent received by the WWTP, but increased to a maximum of 13.73% BD, however, decreased to 10.19% in the FE. A 5.22% difference in the relative abundance of the 120-bp T-RF was observed between the DS and US samples. In contract, the 119-bp T-RF with a relative abundance of 15.26% from the RI received by the WWTP decreased rapidly to undetectable levels BD and then increased to 2.02% in the FE. The relative abundance of the 119-bp T-RF in US sample was observed to be a maximum of 13.51%, whilst it decreased to 4.26% DS. The 36-, 100-, 201-, 204- and 486-bp T-RFs also varied in relative abundances with maximum relative abundances of 14.95%, 22.20%, 16.10%, 21.97%, and 13.42%, respectively, and minimum of zero for all T-RFs. The T-RFs of 80- and 429-bp were less variable remaining below 2.5% across all sample sites, whilst the T-RF for 118-bp remained below 5% across all sample sites. The remaining 47 T-RFs showed undetectable levels across three or more sample sites, with varying relative abundance from 0% to 24.47% (Figure 5.1a).

For Plant B relative abundances of 120-, 204-, 366-, 484-, and 486-bp T-RFs varied greatly (Figure 5.1b). The 484-bp T-RF was present at a relative abundance of 7.45% in the RI received by the WWTP, but increased to a maximum of 29.24% BD, before decreasing rapidly to undetectable levels in FE. The relative abundance of the 484-bp T-RF US was observed to be 2.93%, whilst decreasing to undetectable levels in DS sample. In contrast, the 204-bp T-RF with a relative abundance of 9.76% from the RI received by the WWTP decreased rapidly to undetectable levels upon treatment in BD sample and then increased to a maximum of 56.35% in FE. The relative abundance of the T-RF US was observed to be a maximum of 13.77%, whilst DS showed an increased 51.42%. The 120-, 366- and 486-bp T-RF also varied in relative abundances with maximum relative abundances of 27.74%, 8.25%, and 13.38%, respectively,

and minimum of zero for all T-RFs. The remaining 20 T-RFs showed undetectable levels across three or more sample sites, with varying relative abundance from 0% to 43.89% (Figure 5.1b).

To assess the interspecies abundance ratios of different bacterial species, Pareto-Lorenz curve distribution patterns of the consensus T-RFLP profiles were plotted based on the numbers of peaks and their relative abundance (Figure 5.2). In Plant A, it was observed that 20% of the peaks (number based) corresponded with 81, 81, 75, 79 and 82% of the cumulative relative abundance of the peaks for RI, BD, FE, US and DS, respectively (Figure 5.2a). For Plant B, it was observed that 20% of the peaks (number based) corresponded with 79, 77, 84, 63 and 81% of the cumulative relative abundance of the peaks for RI, BD, FE, US and DS, respectively (Figure 5.2b). Hence, for majority of the T-RFLP profiles, with the exception of US from Plant B, only a small group of bacterial species was numerically dominant and all the others were present in low numbers, with a large difference between the two groups.

5.4.2 Diversity indices and nonmetric multidimensional scaling analysis

Shannon's diversity index (H'), Evenness index (J'), Simpson's diversity index (D) and Simpson's reciprocal diversity index ($1/D$) for each WWTP and the receiving rivers are shown in Figure 5.3. Although these indices are not a definitive measure of diversity, they still provide relative values for assessing spatial variation among different sample sites (Hewson et al. 2007).

Using these indices, the highest diversity for Plant A was observed in the FE sample with values of 3.09, 0.88, 0.93, 15.30 for H', J', D and $1/D$, respectively. The lowest diversity was observed

from the BD sample with values of 2.50, 0.80, 0.88, 8.85 for *H′*, *J′*, *D* and *1/D*, respectively (Figure 5.3a). Significant differences ($p<0.05$) were observed when comparing the indices observed between each sample site. The highest diversity for Plant B was observed US of the receiving river with values of 3.03, 0.90, 0.94, 16.62 for *H′*, *J′*, *D* and *1/D*, respectively, while the lowest diversity was observed in the FE sample with values of 1.57, 0.59, 0.67, 3.08 for *H′*, *J′*, *D* and *1/D*, respectively (Figure 5.3b). Comparison of the indices between each sample sites showed highly significant differences ($p<0.01$) in the indices observed.

Nonmetric multidimensional scaling (NMDS) analysis was done to compare overall similarities and differences in bacterial populations between the five sites based on the T-RFLP profiles in the WWTPs. Distances between points in the NMDS ordination as seen in Figure 5.4 reflects relative dissimilarity of bacterial populations between samples based on T-RFLP profile. For both plants, all replicates of the sample sites clustered closely to each other as expected. For Plant A, DS, BD, FE and US profiles cluster more closely compared to RI profiles (Figure 5.4a). Kruskal's stress values were lower than 0.05 for Plant A, suggesting a good representation of the relationship between the points in the matrix and provide an excellent representation in the reduced dimensions. However, looking at the dendrogram plot of the Bray-Curtis coefficients between the profiles (Figure 5.5a), RI and DS profiles clustered closely compared to BD, FE and US profiles, with FE and US profiles clustering more closely than with BD profiles. For Plant B, all sample site replicate profiles cluster closely to each other compared to the sample sites as expected (Figure 5.4b). Kruskal's stress values were greater than 0.1, but less than 0.2, suggesting a fair representation of the relationship between the points in the matrix and provide a good representation in the reduced dimensions. Looking at the dendrogram plot of the Bray-Curtis coefficients between the profiles (Figure 5.5b), RI

and US clustered closely compared to BD, FE and DS, with FE and DS clustering more closely than to BD profiles.

The ANOSIM reports include R-statistics and p-values. A R-statistic close to 1 indicates that samples in the same group are more similar to each other than samples in different groups. A R-statistic close to 0 indicates that samples in the same group are not more similar to each other than samples in different groups. The *p* value reflects the statistical significance of the R-statistic. ANOSIM of the statistical significance of differences among the T-RFs by the different sample sites showed that for all samples from the two different WWTPs, T-RFs in the triplicate profiles of the sample were highly significantly similar compared to the various sample sites ($R=1$, $p<0.01$).

5.5 Discussion

Understanding the diversity of the aquatic microorganisms is essential for sustainable management of water sources. In addition, changes in bacterial community structures can potentially be used as bio-indicators of environmental disturbances (Ager et al. 2010). In this study, bacterial community diversity and composition were examined by T-RFLP fingerprinting technique. The technique is a highly reproducible, robust, inexpensive and simple tool that has been reported to be useful in characterization of microbial community structures in different habits (Osborn et al. 2000; Prakash et al. 2014). T-RFLP is a method which investigates variations in the lengths of T-RFs of conserved molecular markers, such as

the case in this study which utilised the 16S rRNA gene found in all bacterial species. Although the technique has great potential in characterization of microbial communities, due to the inherent technical biases from PCR as well as instrumental artefacts during peaks detection, meticulous care must be taken when interpreting results from T-RFLP profiles (Danovaro et al. 2006; Cotton et al. 2014). Also, the sequences of 16S rRNA genes have been shown to be very similar among different bacterial species and, thus, are not completely effective in distinguishing among closely related genomic groups due to the generation of T-RF with the same size (Danovaro et al. 2006; Wang et al. 2010). However, in order to overcome the technical artefacts resulting during peak detection and subsequent data analysis, guidelines suggested by Fredriksson et al. (2014b) were applied for all T-RFLP profiles. Furthermore, the increasingly popular trend in T-RFLP analysis has been to associate the T-RFs determined with representative microorganisms using *in silico* digestion of databases with known 16S rRNA genes or the use of T-RFLP clones. However, several studies have reported discrepancies ranging from one to seven base-pairs between *in silico*-determined T-RF length and the actual T-RF length determined by clone T-RFLP (Kitts 2001; Kaplan et al. 2001; Kaplan and Kitts 2003). This can greatly alter the identification of bacterial species of a particular community in question and hence, in this study, *in silico* digestion was not done.

5.5.1 Plant A and suburban river body

Overall, there was a difference in the composition, evenness and diversity indices of the bacterial communities observed across all sample sites. This was as expected as the effects of various treatment stages of the wastewater in Plant A could have resulted in changes in the bacterial community composition and diversity as previously reported (Ye and Zhang 2013).

Surprisingly, the bacterial composition was higher in the final effluent discharge than the influent received by the WWTP suggesting that the treatment was not efficient in reducing bacteria. Furthermore, composition and diversity was observed to be highest at the FE point compared to RI and BD sites of the WWTP as well as the US and DS points of the suburban river. Analysis of the bacterial communities with more sample sites during the various process and the physicochemical parameters of the wastewater might assist in confirming whether or not the increase in the bacterial richness and diversity in the final effluent discharged is a result of the ineffectiveness of the treatment process. However, assessment of the efficiency of the treatment processes was beyond the scope of this study.

The increase in bacterial composition in the final effluent discharged could be due to the possible selection for resistant bacterial species following chlorination and subsequent proliferation of those species due to the presence of dissolved organic nutrients. Although, disinfection treatment by chlorination of the wastewater is expected to reduce the bacterial load, simultaneous protective functions provided by antibiotic resistance genes for chlorine resistance in resistant strains has been previously reported (Shi et al. 2013; Yuan et al. 2015). For example, Jai et al. (2015) reported residual chlorine as the key contributing factor to the bacterial community shift in selection for resistant bacteria with antibiotic resistance genes which allows for chlorine resistance (Jia et al. 2015). This was also observed in another study where the use of chlorine disinfection over UV disinfection resulted in the survival and increased abundance of pathogenic *Escherichia coli*, *Salmonella typhimurium*, *Vibrio cholera* as well as somatic and F-RNA coliphages despite high levels of residual chlorine concentrations in the treated effluent (Dungeni et al. 2010). Therefore, it is not surprising to observe the slight increase in bacterial communities downstream of the receiving suburban

river body in terms of composition and diversity compared to the US sample, which could be attributed to the effect of the final effluent discharge. However, it must be mentioned that the results of Plant A in this study contradicts what has been previously reported, where reduction in the abundance and diversity of benthic bacterial communities in urban and suburban rivers was observed (Drury et al. 2013).

5.5.2 Plant B and urban river body

Plant B resulted in significantly altered bacterial community composition and diversity downstream of the receiving river, in contrast to Plant A. The most striking effect was the absence of bacterial species present in upstream samples at the downstream sites, in addition to downstream sites having relatively the same community composition as the wastewater effluent discharged. A potential reason for the decreased levels of composition and diversity for Plant B could be due to the presence and inhibitory effects of toxic compounds in the final treated effluent. Several studies have shown the effects of a wide range of biologically active compounds, including low levels of antimicrobials and dissolved heavy metals, in rivers and streams as a result of effluent dispersed from nearby WWTP (Radjenović et al. 2009; Oulton et al. 2010; Jelic et al. 2011; Gatica and Cytryn 2013; Chirila et al. 2014; Collado et al. 2014). Many of these compounds are not completely removed during wastewater treatment, thus making WWTP a major source for their dissemination into the environment (Ort et al. 2010; Chirila et al. 2014). Furthermore, it has been reported that even at low concentrations in the environment, these compounds can have adverse effects on aquatic organisms (Radjenovic et al. 2007; Chirila et al. 2014). The observed decrease in the bacterial composition and diversity downstream of the receiving rivers in this study could be attributed to the possible presence of

compounds toxic to microorganisms, However, future studies will attempt to quantify the biologically active and toxic compounds in the effluents to corroborate this speculation.

The results obtained in this study correlates with previous finding which also showed a decrease in microbial community richness and diversity phenomenon from effluent discharged into aquatic ecosystems (Drury et al. 2013). The results illustrates the concept of biotic homogenization, which suggests that human modifications of the environment are reducing the biological richness that exist in natural ecosystems and consequently only a subset of naturally occurring species are constantly being selected for as a result of the human-altered ecosystems (McKinney 2008; Baiser et al. 2012). Although, this phenomenon has been largely seen in plant and animal community studies, it has been poorly documented for microbial communities (McKinney 2008). Effects of this phenomenon has been suggested to result in a more homogenized biosphere with lower diversity on regional and global scales (McKinney and Lockwood 1999). Therefore, our results suggests that Plant B's effluent may be a driving force for biotic homogenization of bacterial communities in receiving river and may lead to negative implications on the ecosystem functions.

Although the popularity of T-RFLP seems to be decreasing as the method of choice for community profiling and analysis, due to the increasing popularity, availability and decreasing costs of next generation sequencing, however, the technique still remains a quick, easy and cost effective method for an initial indication of the bacterial community diversity and composition. Our results demonstrate that the two rivers which differed in the levels of urbanization showed different responses to WWTP effluent inputs, including changes in the composition and

diversity of the bacterial communities. Given that majority of the design and intended operation of WWTPs in the KwaZulu-Natal province of South Africa are similar to the remaining provinces, these results raises new questions on the effects of human modification of river ecosystems. Nonetheless, the present study sheds some light on the dynamics of bacterial communities during wastewater treatment in relation to the effects of treated effluent discharge on the river ecosystems. Furthermore, our study further support the concept of biotic homogenization of bacterial communities in receiving river bodies as a result of effluent discharge from WWTPs. Hence, more studies and investigations are still needed to shed some light on this important topic.

5.6 References

Ager D, Evans S, Li H, Lilley AK, van der Gast CJ (2010) Anthropogenic disturbance affects the structure of bacterial communities. Environ Microbiol 12:670–678. doi: 10.1111/j.1462-2920.2009.02107.x

Ahmad I, Ahmad F, Pichtel J (eds) (2011) Microbes and Microbial Technology. Springer New York, New York, NY

Baiser B, Olden JD, Record S, Lockwood JL, McKinney ML (2012) Pattern and process of biotic homogenization in the New Pangaea. Proc Biol Sci 279:4772–4777. doi: 10.1098/rspb.2012.1651

Blackwood CB, Hudleston D, Zak DR, Buyer JS (2007) Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. Appl Environ Microbiol 73:5276–83. doi: 10.1128/AEM.00514-07

Brooks BW, Riley TM, Taylor RD (2006) Water quality of effluent-dominated ecosystems: ecotoxicological, hydrological, and management considerations. Hydrobiologia 556:365–379. doi: 10.1007/s10750-004-0189-7

Carey RO, Migliaccio KW (2009) Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems: a review. Environ Manage 44:205–217. doi: 10.1007/s00267-009-9309-5

Chirila E, Draghici C, Puhacel A (2014) Total and dissolved metals occurrence in municipal wastewater treatment plant effluents. Environ Eng Manag J 13:2211–2218.

Collado N, Rodriguez-Mozaz S, Gros M, Rubirola A, Barceló D, Comas J, Rodriguez-Roda I, Buttiglieri G (2014) Pharmaceuticals occurrence in a WWTP with significant industrial contribution and its input into the river system. Environ Pollut 185:202–212. doi: 10.1016/j.envpol.2013.10.040

Cotton TEA, Dumbrell AJ, Helgason T (2014) What goes in must come out: testing for biases in molecular analysis of arbuscular mycorrhizal fungal communities. PLoS One 9:e109234. doi: 10.1371/journal.pone.0109234

Danovaro R, Luna GM, Dell'anno A, Pietrangeli B (2006) Comparison of two fingerprinting techniques, terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analysis, for determination of bacterial diversity in aquatic environments. Appl Environ Microbiol 72:5982–5989. doi: 10.1128/AEM.01361-06

Department of Water and Sanitation (2015) Green Drop Progress Report. https://www.dwa.gov.za/dir_ws/gds/Docs/DocsDefault.aspx. Accessed 28 Apr 2016

Disayathanoowat T, Young JPW, Helgason T, Chantawannakul P (2012) T-RFLP analysis of bacterial communities in the midguts of *Apis mellifera* and *Apis cerana* honey bees in Thailand. FEMS Microbiol Ecol 79:273–281. doi: 10.1111/j.1574-6941.2011.01216.x

Drury B, Rosi-Marshall E, Kelly JJ (2013) Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 79:1897–1905. doi: 10.1128/AEM.03527-12

Dungeni M, van der Merwe R, Momba M (2010) Abundance of pathogenic bacteria and viral indicators in chlorinated effluents produced by four wastewater treatment plants in the Gauteng Province, South Africa. Water SA 36:607–614.

Fredriksson NJ, Hermansson M, Wilén B-M (2014a) Tools for T-RFLP data analysis using Excel. BMC Bioinformatics 15:361. doi: 10.1186/s12859-014-0361-7

Fredriksson NJ, Hermansson M, Wilén B-M (2014b) Impact of T-RFLP data analysis choices on assessments of microbial community structure and dynamics. BMC Bioinformatics 15:360. doi: 10.1186/s12859-014-0360-8

Gatica J, Cytryn E (2013) Impact of treated wastewater irrigation on antibiotic resistance in the soil microbiome. Environ Sci Pollut Res Int 20:3529–3538. doi: 10.1007/s11356-013-1505-4

Gücker B, Brauns M, Pusch MT (2006) Effects of wastewater treatment plant discharge on ecosystem structure and function of lowland streams. J North Am Benthol Soc 25:313–329. doi: 10.1899/0887-3593(2006)25[313:EOWTPD]2.0.CO;2

Hewson I, Jacobson Meyers ME, Fuhrman JA (2007) Diversity and biogeography of bacterial assemblages in surface sediments across the San Pedro Basin, Southern California Borderlands. Environ Microbiol 9:923–933. doi: 10.1111/j.1462-2920.2006.01214.x

Jelic A, Gros M, Ginebreda A, Cespedes-Sánchez R, Ventura F, Petrovic M, Barcelo D (2011) Occurrence, partition and removal of pharmaceuticals in sewage water and sludge during wastewater treatment. Water Res 45:1165–1176. doi: 10.1016/j.watres.2010.11.010

Jia S, Shi P, Hu Q, Li B, Zhang T, Zhang X-X (2015) Bacterial community shift drives antibiotic resistance promotion during drinking water chlorination. Environ Sci Technol 49:12271–12279. doi: 10.1021/acs.est.5b03521

Kaplan CW, Astaire JC, Sanders ME, Reddy BS, Kitts CL (2001) 16S Ribosomal DNA terminal restriction fragment pattern analysis of bacterial communities in feces of rats fed *Lactobacillus acidophilus* NCFM. Appl Environ Microbiol 67:1935–1939. doi: 10.1128/AEM.67.4.1935-1939.2001

Kaplan CW, Kitts CL (2003) Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. J Microbiol Methods 54:121–125.

Kitts CL (2001) Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr Issues Intest Microbiol 2:17–25.

Lawrence JR, Swerhone GDW, Wassenaar LI, Neu TR (2005) Effects of selected pharmaceuticals on riverine biofilm communities. Can J Microbiol 51:655–669. doi: 10.1139/w05-047

Legendre P, Legendre L (1998) Numerical Ecology, 2nd Englis. Elsevier Science BV, Amsterdam, Netherlands

Marzorati M, Wittebolle L, Boon N, Daffonchio D, Verstraete W (2008) How to get more out of molecular fingerprints: practical tools for microbial ecology. Environ Microbiol 10:1571–1581. doi: 10.1111/j.1462-2920.2008.01572.x

McKinney M, Lockwood J (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol Evol 14:450–453.

McKinney ML (2008) Effects of urbanization on species richness: A review of plants and animals. Urban Ecosyst 11:161–176. doi: 10.1007/s11252-007-0045-4

Ort C, Lawrence MG, Rieckermann J, Joss A (2010) Sampling for pharmaceuticals and personal care products (PPCPs) and illicit drugs in wastewater systems: are your conclusions valid? A critical review. Environ Sci Technol 44:6024–6035. doi: 10.1021/es100779n

Osborn AM, Moore ER, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ Microbiol 2:39–50.

Oulton RL, Kohn T, Cwiertny DM (2010) Pharmaceuticals and personal care products in effluent matrices: A survey of transformation and removal during wastewater treatment and implications for wastewater management. J Environ Monit 12:1956–1978. doi: 10.1039/c0em00068j

Possemiers S, Verthé K, Uyttendaele S, Verstraete W (2004) PCR-DGGE-based quantification of stability of the microbial community in a simulator of the human intestinal microbial ecosystem. FEMS Microbiol Ecol 49:495–507. doi: 10.1016/j.femsec.2004.05.002

Prakash O, Pandey PK, Kulkarni GJ, Mahale KN, Shouche YS (2014) Technicalities and glitches of terminal restriction fragment length polymorphism (T-RFLP). Indian J Microbiol 54:255–61. doi: 10.1007/s12088-014-0461-0

R Development Core Team (2007) R: a language and environment for statistical computing.

Radjenovic J, Petrovic M, Barceló D (2007) Analysis of pharmaceuticals in wastewater and removal using a membrane bioreactor. Anal Bioanal Chem 387:1365–1377. doi: 10.1007/s00216-006-0883-6

Radjenović J, Petrović M, Barceló D (2009) Fate and distribution of pharmaceuticals in wastewater and sewage sludge of the conventional activated sludge (CAS) and advanced membrane bioreactor (MBR) treatment. Water Res 43:831–841. doi: 10.1016/j.watres.2008.11.043

Sarti A, Pozzi E, Zaiat M (2012) Characterization of immobilized biomass by amplified rDNA restriction analysis (ARDRA) in an anaerobic sequencing-batch biofilm reactor (ASBBR) for the treatment of industrial wastewater. Brazilian Arch Biol Technol 55:623–629. doi: 10.1590/S1516-89132012000400019

Shen SY, Fulthorpe R (2015) Seasonal variation of bacterial endophytes in urban trees. Front Microbiol 6:427. doi: 10.3389/fmicb.2015.00427

Shi P, Jia S, Zhang X-X, Zhang T, Cheng S, Li A (2013) Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res 47:111–120. doi: 10.1016/j.watres.2012.09.046

Tipayno S, Kim C-G, Sa T (2012) T-RFLP analysis of structural changes in soil bacterial communities in response to metal and metalloid contamination and initial phytoremediation. Appl Soil Ecol 61:137–146. doi: 10.1016/j.apsoil.2012.06.001

Vartoukian SR, Palmer RM, Wade WG (2010) Strategies for culture of "unculturable" bacteria. FEMS Microbiol Lett 309:1–7. doi: 10.1111/j.1574-6968.2010.02000.x

Wakelin SA, Colloff MJ, Kookana RS (2008) Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. Appl Environ Microbiol 74:2659–2668. doi: 10.1128/AEM.02348-07

Wang X, Wen X, Criddle C, Yan H, Zhang Y, Ding K (2010) Bacterial community dynamics in two full-scale wastewater treatment systems with functional stability. J Appl Microbiol 109:1218–1226. doi: 10.1111/j.1365-2672.2010.04742.x

WHO (2006) Guidelines for the safe use of wastewater, excreta and greywater - Policy and regulatory aspects. http://www.who.int/water_sanitation_health/wastewater/gsuww/en/. Accessed 28 Apr 2016

Ye L, Zhang T (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Appl Microbiol Biotechnol 97:2681–2690. doi: 10.1007/s00253-012-4082-4

Yuan Q-B, Guo M-T, Yang J (2015) Fate of antibiotic resistant bacteria and genes during wastewater chlorination: implication for antibiotic resistance control. PLoS One 10:e0119403. doi: 10.1371/journal.pone.0119403

5.7 Figures and tables

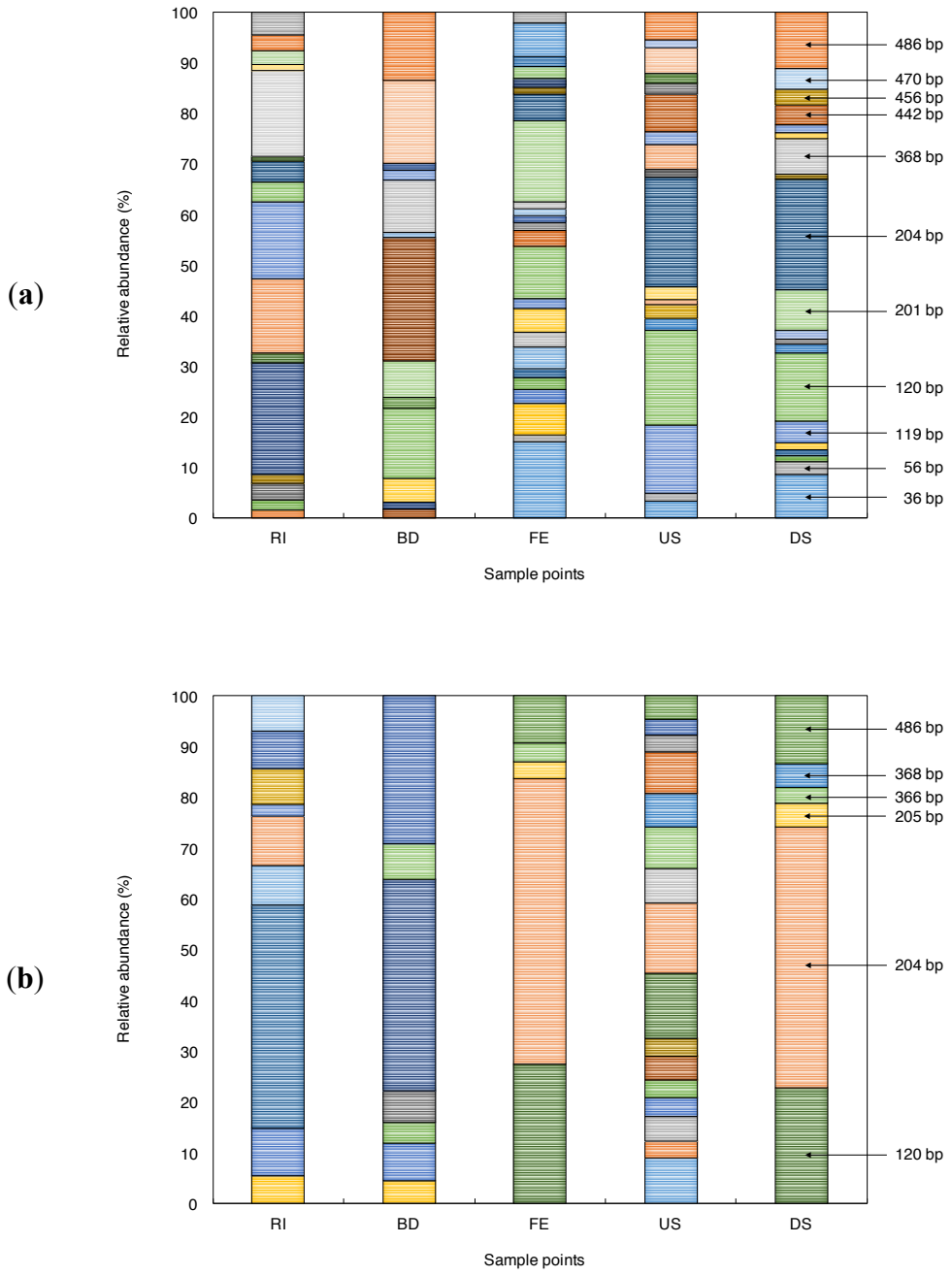
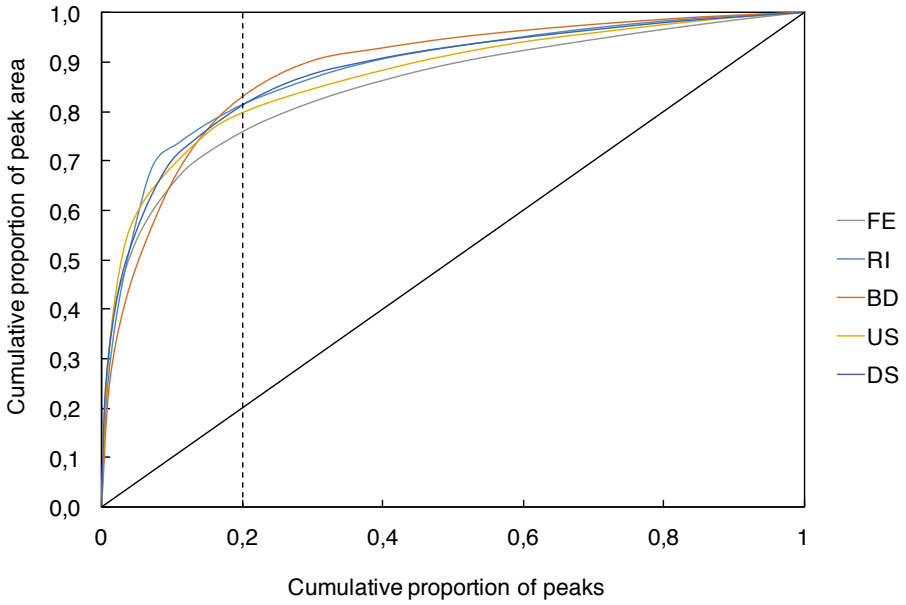


Figure 5.1: Histograms of terminal-restriction fragments (T-RF) relative abundances in (**a**) Plant A and (**b**) Plant B. Each T-RF is the mean (n=3) size upon normalization, alignment, alignment correction, systematic differences detection. The relative abundance is the ratio of the peak area of a given T-RF in a given sample to the sum of all T-RFs in that sample expressed as a percentage. Arrows indicate the consensus sizes from three replicates (upon normalization, alignment, alignment correction, systematic differences detection) of the restriction fragments for dominant T-RF in base pairs (bp).

(a)



(b)

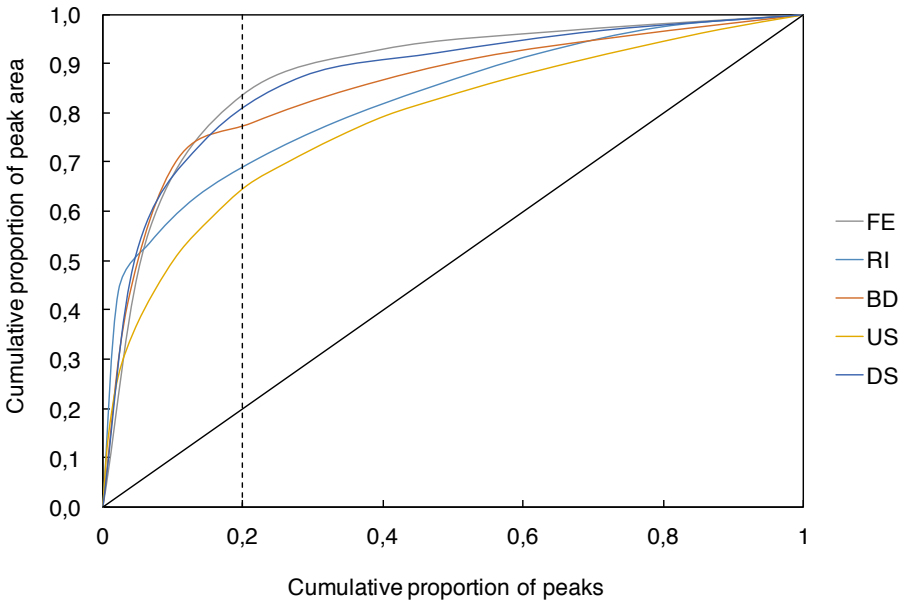


Figure 5.2: Pareto-Lorenz distribution curves of the average bacterial terminal-restriction fragment length polymorphism profiles from the various sampling points in this study of (**a**) Plant A and (**b**) Plant B. Each T-RF is the mean (*n*=3) size upon normalization, alignment, alignment correction, systematic differences detection. The dashed vertical line at the 0.2 x-axis level is plotted to evaluate the range of the Pareto values, whilst the 45° diagonal line is the theoretical line representing perfect evenness.

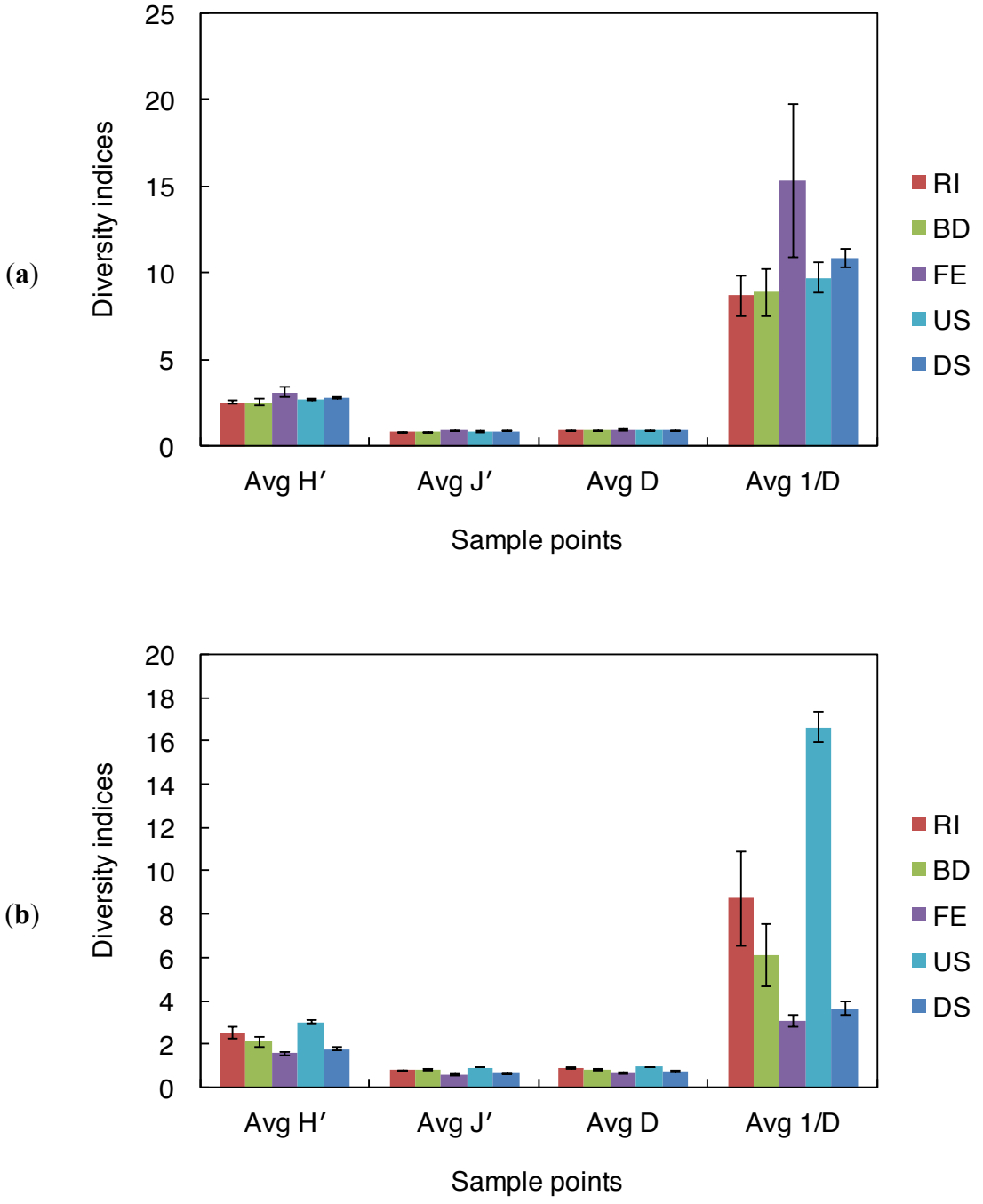
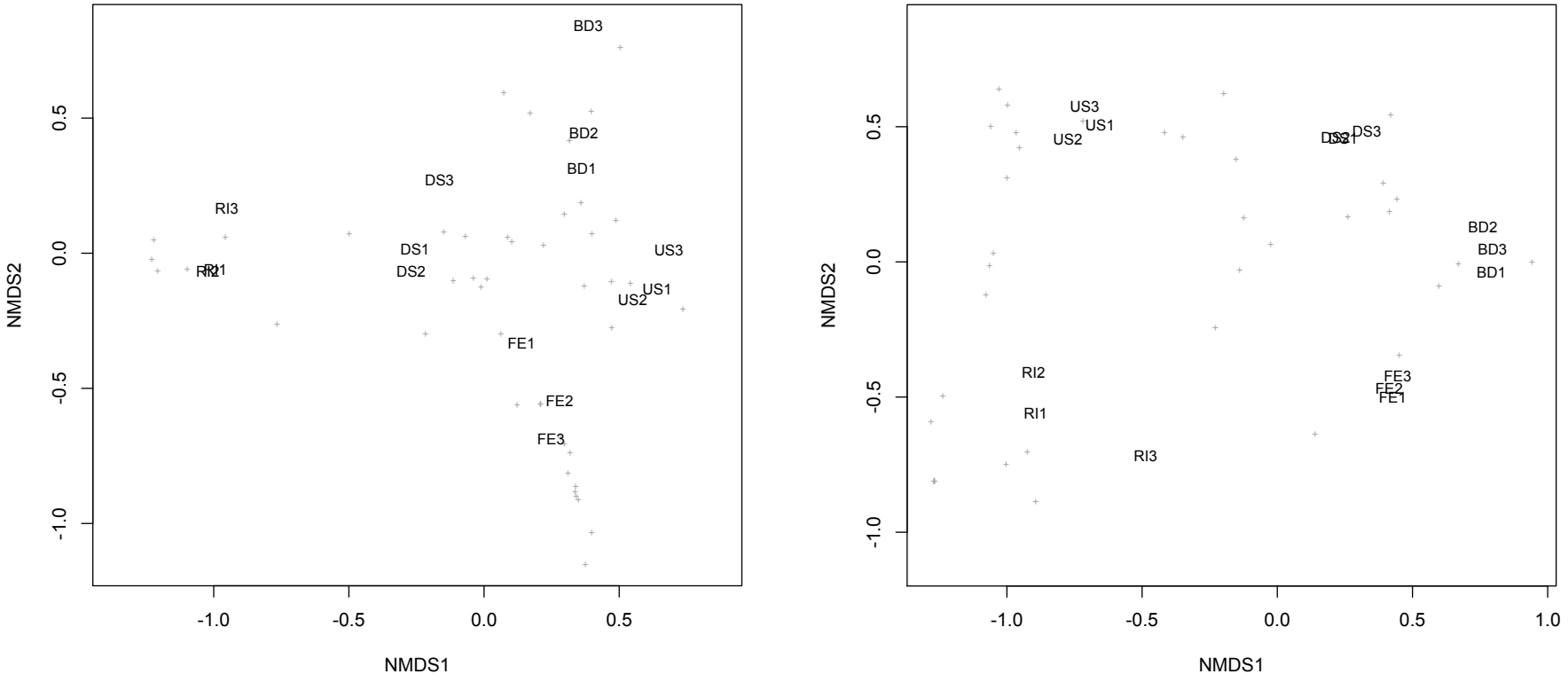


Figure 5.3: Mean (average) diversity indices determined for (**a**) Plant A (**b**) Plant B from the various sampling points in this study: Shannon's diversity index (*H′*), Evenness index (*J′*), Simpson's diversity index (*D*) and Simpson's reciprocal diversity index (*1/D*). Each data point is the mean (n=3) ± standard error. One-way analysis of variance (ANOVA) of the indices demonstrated significant differences (p<0.05) between each sample sites for Plant A and highly significant differences (p<0.01) between each sample sites for Plant B.



(**a**) (**b**)

Figure 5.4: Nonmetric multidimensional scaling (NMDS) plot for the first two dimensions of the Bray-Curtis dissimilarity coefficients of the bacterial community terminal restriction fragment length polymorphism (T-RFLP) profiles from the various sampling points in this study of (**a**) Plant A, stress = 0.09839739 and (**b**) Plant B, stress = 0.1145624. Grey cross represents the unique relative abundance of terminal-restriction fragments in each T-RFLP profile. Profile sample sites with the replicate number represents the T-RFLP profile ordination. Abbreviations: RI, raw influent; BD, wastewater before disinfection stage; FE, final effluent discharged; US, upstream of final effluent discharge site on the river; DS, downstream of final effluent discharge site on the river.

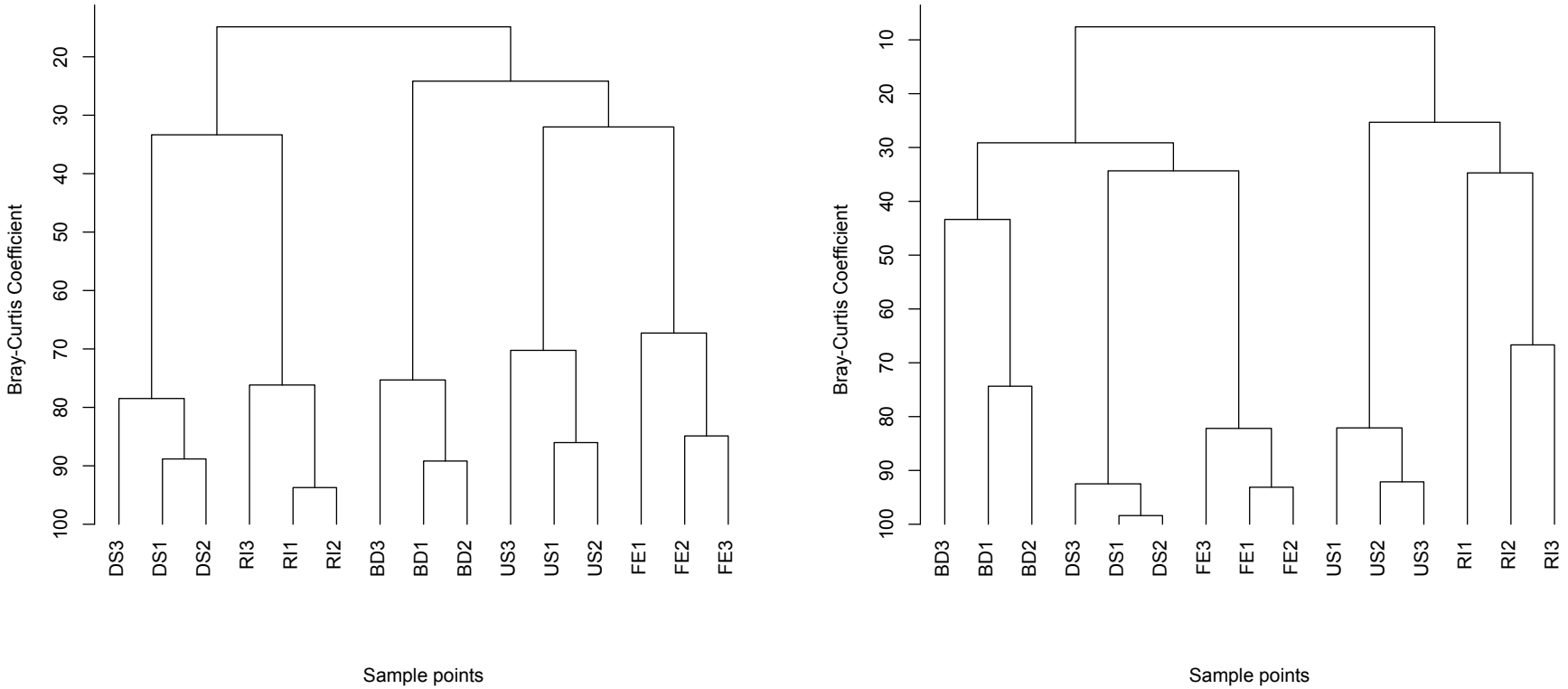


Figure 5.5: Dendrogram plot of Bray-Curtis dissimilarity coefficients of the bacterial community T-RFLP profiles from the various sampling points in this study of (**a**) Plant A (**b**) Plant B. Abbreviations: RI, raw influent; BD, wastewater before disinfection stage; FE, final effluent discharged; US, upstream of final effluent discharge site on the river; DS, downstream of final effluent discharge site on the river.

Table 5.1: Mean terminal-restriction fragments (T-RF) richness (*S*) of the various sampling points in this study from Plant A and Plant B. Each T-RF is the mean (*n*=3) size upon normalization, alignment, alignment correction, systematic differences detection.

	RI	**BD**	**FE**	**US**	**DS**
Plant A	16	13	24	18	20
Plant B	9	7	5	16	6

Chapter Six

Manuscripts formatted and intended for submission to *Genome Announcement*

6.1 Draft genome sequence of an *Acinetobacter* sp. PT1 reconstructed from metagenomic sequences of wastewater and river samples in Durban, South Africa

6.1.1 Abstract

*Acinetobacter s*p. have been reported to be ubiquitous in the environment, many of which may be pathogenic and/or express multidrug resistance. Here, we report the 2.324 Mb draft genome sequence of the *Acinetobacter* sp. PT1, reconstructed from the metagenomes of wastewater and river samples collected in the city of Durban, South Africa.

6.1.2 Genome announcement

Acinetobacter is a complex genus which includes Gram-negative, non-motile, aerobic coccobacilli bacteria that are able to inhabit several ecological niches including soil, water, animals and humans (1). Alarmingly, *A. baumannii* have been associated with nosocomial infections, with increasing number of multi-drug resistant strains resulting in high rates of mortality and long periods of hospitalization (2, 3). Here, we present a draft genome of an *Acinetobacter* sp. PT1 which was obtained from multiple metagenomic sequences of wastewater samples collected at various treatment stages from a wastewater treatment plant

(WWTP) in Durban, South Africa as well as samples from the river receiving treated effluent discharge from this WWTP.

Total DNA was extracted from each of the wastewater samples using the Power Water DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Using a whole-genome shotgun strategy, the metagenomic libraries were prepared using the GS FLX Titanium Rapid Library Preparation Kit (Roche, USA) and sequenced with a Roche 454 genome sequencer (GS FLX titanium) pyrosequencing system at the National Genomics Platform (Technology Innovation Agency, South Africa). *De novo* cross assembly was carried out using IDBA_UD (4) and subsequently, bins were generated from binning of the assembled contigs based on metagenomic read coverage, tetranucleotide frequency and the occurrence of unique marker genes using MaxBin (5). CheckM (6) estimated the completeness of the draft genome to be 75% with 3% contamination. Our search of the contigs bin against all bacterial and archaeal genome sequences available in the GenBank database (February 2016) using BLASTn (7) showed best hits for *Acinetobacter* spp.

The draft genome is 2,324,032 bp with 1217 contigs (ranging from 1 to 7.9 kb) and an average GC content of 42% (with N_{50} value of 2011). Gene prediction and functional annotation of the draft genome performed using the RAST server (8) placed PT1 to share an ancestral node with the *A. johnsonii* cluster which further shared an ancestor with the *A. lwoffii*, *A. schindleri*, *Acinetobacter* sp. HA cluster based on 1327 and 1513 different cluster signatures, respectively, suggesting that PT1 is closely related to *A. johnsonii*. Additionally, RAST annotated 35 RNA genes, whilst 2250 protein-coding sequences were assigned to 278 SEED subsystem categories.

In addition to the expected genes encoding catabolism, biosynthesis, and stress response, the draft genome was annotated with genes encoding prophages, transposable elements and plasmids, and 15 genes under virulence, disease and defense category. ARG-ANNOT (9) identified several resistance genes encoding β-lactamases ($bla_{OXA-211}$, $bla_{OXA-212}$ and $bla_{OXA-309}$), macrolide resistance (*mphE* and *msrE*), and tetracycline resistance determinant (*tet39*), which have been reported to be found in several *Acinetobacter* species (10).

Nucleotide sequence accession number(s)

The sequences from the whole-genome shotgun project investigating *Acinetobacter* sp. PT1 have been deposited in DDBJ/EMBL/GenBank and is currently under review with submission no. SUB1581611 and awaiting designation of accession and version no.

6.1.3 References

1. **Warskow AL**, **Juni E**. 1972. Nutritional requirements of *Acinetobacter* strains isolated from soil, water, and sewage. J. Bacteriol. **112**:1014–1016.
2. **Falagas ME**, **Bliziotis IA**, **Siempos II**. 2006. Attributable mortality of *Acinetobacter baumannii* infections in critically ill patients: a systematic review of matched cohort and case-control studies. Crit. Care **10**:R48.
3. **Dijkshoorn L**, **Nemec A**, **Seifert H**. 2007. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. Nat. Rev. Microbiol. **5**:939–951.

4. **Peng Y**, **Leung HCM**, **Yiu SM**, **Chin FYL**. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28**:1420–1428.

5. **Wu Y-W**, **Tang Y-H**, **Tringe SG**, **Simmons BA**, **Singer SW**. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome **2**:26.

6. **Parks DH**, **Imelfort M**, **Skennerton CT**, **Hugenholtz P**, **Tyson GW**. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. PeerJ Prepr. **2**:e554v1 doi:.10.7287/peerj.preprints.554v1.

7. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

8. **Aziz RK**, **Bartels D**, **Best AA**, **DeJongh M**, **Disz T**, **Edwards RA**, **Formsma K**, **Gerdes S**, **Glass EM**, **Kubal M**, **Meyer F**, **Olsen GJ**, **Olson R**, **Osterman AL**, **Overbeek RA**, **McNeil LK**, **Paarmann D**, **Paczian T**, **Parrello B**, **Pusch GD**, **Reich C**, **Stevens R**, **Vassieva O**, **Vonstein V**, **Wilke A**, **Zagnitko O**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

9. **Gupta SK**, **Padmanabhan BR**, **Diene SM**, **Lopez-Rojas R**, **Kempf M**, **Landraud L**, **Rolain J-M**. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob. Agents Chemother. **58**:212–220.

10. **Figueiredo S**, **Bonnin RA**, **Poirel L**, **Duranteau J**, **Nordmann P**. 2012. Identification of the naturally occurring genes encoding carbapenem-hydrolysing oxacillinases from *Acinetobacter haemolyticus*, *Acinetobacter johnsonii*, and *Acinetobacter calcoaceticus*. Clin. Microbiol. Infect. **18**:907–913.

6.2 Draft genome sequence of a *Polynucleobacter necessarius* PT2 reconstructed from metagenomic sequences of wastewater and river samples in Durban, South Africa

6.2.1 Abstract

Polynucleobacter necessarius spp. have been reported to be ubiquitous in freshwater environment, many of which have been reported to be endosymbiont of the ciliate protist *Euplotes aediculatus*. Here, we report the 1.414 Mb draft genome sequence of the *P. necessarius* PT2, reconstructed from the metagenomes of wastewater and river samples collected in the city Durban, South Africa

6.2.2 Genome announcement

The *Polynucleobacter* genus falls under the Proteobacteria phylum, under the Beta-proteobacteria class and is affiliated with the Burkholderiaceae family. The genus, which harbors Gram-negative, non-motile, rod-shaped bacteria was originally established by Heckmann and Schmidt in 1987 (1). Organisms from this genus have been reported to be cytoplasmic endosymbionts of the ciliate protist *Euplotes aediculatus*. To date, there are only four members under this genus *viz. P. necessarius*, *P. cosmopolitanus*, *P. rarus* and *P. acidiphobus*, and by definition, the genus would have been designated as Candidatus (2). Here,

we present a draft genome of a *Polynucleobacter necessarius* PT2 which was obtained from multiple metagenomic sequences of wastewater samples collected at various treatment stages from a wastewater treatment plant (WWTP) in Durban, South Africa as well as samples from the river receiving treated effluent discharge from this WWTP.

Total DNA was extracted from each of the wastewater samples using the Power Water DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Using a whole-genome shotgun strategy, the metagenomic libraries was prepared using the GS FLX Titanium Rapid Library Preparation Kit (Roche, USA) and sequenced with a Roche 454 genome sequencer (GS FLX titanium) pyrosequencing system at the National Genomics Platform (Technology Innovation Agency, South Africa). *De novo* cross assembly was carried out using IDBA_UD (3) and subsequently, bins were generated from binning of the assembled contigs based on metagenomic read coverage, tetranucleotide frequency and the occurrence of unique marker genes using MaxBin (4). CheckM (5) estimated the completeness of the draft genome to be 68% with 3% contamination. Our search of the contigs bin against all the bacterial and archaeal genome sequences available in the GenBank database (February 2016) using BLASTn (6) showed best hits for *Polynucleobacter necessarius*.

The draft genome is 1,414,206 bp with 729 contigs (ranging from 1 to 11.5 kb) and an average GC content of 45.6 % (with N_{50} value of 2045). Gene prediction and functional annotation of the draft genome performed using the RAST server (7) placed PT2 to be related with the 1,560,469 bp *P. necessarius* subsp. *necessarius* STIR1 (accession no. NC_010531) (8), which shared an ancestral node with the 2,159,490 bp *P. necessarius* subsp. *asymbioticus* QLW-

P1DMWA-1^T (accession no. NC_009379) (2), based on 859 different cluster signatures, suggesting that PT2 is closely related to *P. necessarius* subsp. *necessarius*. Additionally, RAST annotated 41 RNA genes, whilst 1421 protein-coding sequences were assigned to 219 SEED subsystem categories.

In summary, we present the draft genomic sequence of PT2 which is closely related to the sequence of *P. necessarius* species. As the organisms under this genus are known to be both symbiotic and free-living, further analysis and comparisons of the draft genome will allow for exploration and new insights into the metabolism of the oligotrophic bacteria and their ubiquitous ecological roles in freshwater systems.

Nucleotide sequence accession number(s)

The sequences from the whole-genome shotgun project investigating *Polynucleobacter necessaries* PT2 have been deposited in DDBJ/EMBL/GenBank and is currently under review with submission no. SUB1581622 and awaiting designation of accession and version no.

6.2.3 References

1. **Heckmann K**, **Schmidt HJ**. 1987. *Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. Int. J. Syst. Bacteriol. **37**:456–457.

2. **Meincke L**, **Copeland A**, **Lapidus A**, **Lucas S**, **Berry KW**, **Del Rio TG**, **Hammon N**, **Dalin E**, **Tice H**, **Pitluck S**, **Richardson P**, **Bruce D**, **Goodwin L**, **Han C**, **Tapia R**, **Detter JC**, **Schmutz J**, **Brettin T**, **Larimer F**, **Land M**, **Hauser L**, **Kyrpides NC**, **Ivanova N**, **Göker M**, **Woyke T**, **Wu QL**, **Pöckl M**, **Hahn MW**, **Klenk H-P**. 2012. Complete genome sequence of *Polynucleobacter necessarius* subsp. asymbioticus type strain (QLW-P1DMWA-1(T)). Stand. Genomic Sci. **6**:74–83.

3. **Peng Y**, **Leung HCM**, **Yiu SM**, **Chin FYL**. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28**:1420–1428.

4. **Wu Y-W**, **Tang Y-H**, **Tringe SG**, **Simmons BA**, **Singer SW**. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome **2**:26.

5. **Parks DH**, **Imelfort M**, **Skennerton CT**, **Hugenholtz P**, **Tyson GW**. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. PeerJ Prepr. **2**:e554v1 doi:.10.7287/peerj.preprints.554v1.

6. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

7. **Aziz RK**, **Bartels D**, **Best AA**, **DeJongh M**, **Disz T**, **Edwards RA**, **Formsma K**, **Gerdes S**, **Glass EM**, **Kubal M**, **Meyer F**, **Olsen GJ**, **Olson R**, **Osterman AL**, **Overbeek RA**, **McNeil LK**, **Paarmann D**, **Paczian T**, **Parrello B**, **Pusch GD**, **Reich C**, **Stevens R**, **Vassieva O**, **Vonstein V**, **Wilke A**, **Zagnitko O**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

8. **Boscaro V**, **Felletti M**, **Vannini C**, **Ackerman MS**, **Chain PSG**, **Malfatti S**, **Vergez LM**, **Shin M**, **Doak TG**, **Lynch M**, **Petroni G**. 2013. *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. **110**:18590–18595.

Chapter Seven

7 General discussion, future developments and conclusion

7.1 Research in perspective

The magnitude of the global freshwater crisis is underestimated. Clean and safe drinking water is indispensable for sustenance, health and dignity of life, whilst freshwater is central to energy and food security as well as all ecosystems functions. Continued development of our societies profoundly depends on this natural resource. Though international and national efforts have illuminated this crisis and certain strategies implemented to safeguard these precious resources over the past few decades, continued deterioration of water quality is still a major area of concern. This is owing to the continued reduction of the available options of water usage (World Economic Forum, 2015). Contamination of surface water resources due to discharge of inadequately treated wastewaters has previously been indicated as an important topic requiring continued investigations. Often, these discharges introduce large amounts of organic matter and nutrients which could lead to eutrophication and temporary oxygen deficiencies, ultimately disrupting the natural biotic community structure and its important ecological functions. In addition, pathogenic microorganisms present in the discharged wastewater increases public health risks due to their disease causing potential (Corcoran *et al.*, 2010). This could result in increased cases of diseases such as dysentery, cholera, skin infections and typhoid fever in a country like South Africa with an already stressed public health system. This is further compounded by the increasing cases of multidrug resistant organisms, with consequent increase in morbidity and mortality (Department of Environmental Affairs, 2016). It is

therefore imperative that continued monitoring of wastewater treatment plants (WWTPs) and its treated effluent being discharged into the environment is conducted to gain important understanding of the public health implications to the population serviced. This will allow for implementation of sustainable wastewater treatment strategies, thereby safeguarding the limited freshwater resources (Department of Environmental Affairs, 2016).

Hence, the main focus of this study was to investigate the role of two urban wastewater treatment plants in the dissemination of antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARGs) into the receiving aquatic systems, as well as the general impacts of effluent discharge on the aquatic ecosystem functioning through the use of next-generation sequencing coupled with metagenomic approach.

Results of the bacterial community composition and diversity analysis presented in Chapters two, three and four revealed changes in the community composition upon treatment of the raw wastewaters received by both WWTPs. Members of the Proteobacteria phyla was observed to dominate the bacterial communities in effluents of both WWTPs as well as in both urban and suburban effluent-receiving rivers investigated in this study. Dominance by the Proteobacteria phyla has been reported in previous studies investigating the microbial communities of water treatment plants, such as targeting the 16S rRNA gene of samples collected from WWTPs (Miura *et al*., 2007; Silva *et al*., 2010), DNA cloning of activated sludge (Snaidr *et al.*, 1997), microarray analysis of biological wastewater treatment reactors (Xia *et al.*, 2010), metagenomic analysis of different WWTPs (Hu *et al.*, 2012) and even drinking water treatment plants (Shi *et al.*, 2013). Furthermore, previous studies investigating different ecosystems such

as artic river sources (Kolmakova *et al.*, 2014), soil (Roesch *et al.*, 2007) and freshwater sources (Ghai *et al.*, 2011) also saw dominance of the Proteobacteria phylum. Predominance of this phylum could be explained by the fact that it comprises of one of the most phylogenetically and metabolically versatile group in the Bacteria domain (Ettema and Andersson, 2009) and is known to typically occupy an average of 40% of a given bacterial population (Nemergut *et al.*, 2011). Hence, as WWTPs and rivers present with an ever changing environment, such as the composition and concentration of nutrients and/or pollutants, the observation of dominance by this phylum across all samples is not surprising. An overall decrease in the relative abundance of Proteobacteria from 51% (in raw influent) to 44% (in treated effluent sample) was observed in Plant A. In contrast, Plant B saw an overall increase in the relative abundances of Proteobacteria from 45% (in raw influent) to 84% (in treated effluent sample). In Plant A, a 15% decrease in the relative abundance of members of the Proteobacteria phylum from upstream (80%) to downstream (69%) of the effluent-receiving (suburban) river was observed. Conversely, relative abundances of members of the Proteobacteria phylum saw a 36% increase from the natural baseline population in the river receiving Plant B effluent discharges, shifting from 61% upstream to 83% downstream of the urban river. These differences could be attributed to geographical environmental factors and overall ineffectiveness of the treatment process adopted by the different WWTP. Besides dominance by Proteobacteria, high abundance of several other phyla was seen in both WWTPs as well as in both urban and suburban rivers receiving the treated effluent from these plants, including members belonging to the Bacteroidetes, Actinobacteria, Verrucomicrobia and Firmicutes phyla. Detection of Bacteroidetes, Actinobacteria and Firmicutes has been well documented in recent studies investigating WWTPs, such as untreated wastewaters (Shanks *et al.*, 2013), swine WWTP (Da Silva *et al.*, 2015), anaerobic reactor digesting activated sludge from WWTP (Guo *et al.*, 2015), tannery WWTP (Wang *et al.*, 2013) and activated sludge of a WWTP in Hong Kong (Yu and

Zhang, 2012). Additionally, members of these phyla have also been reported in studies investigating river (Jordaan and Bezuidenhout, 2013; Kolmakova *et al.*, 2014; Sánchez-Andrea *et al.*, 2011; Wu *et al.*, 2012), lake sediments (Sauvain *et al.*, 2013), mangrove (Andreote *et al.*, 2012) and soil (Foong *et al.*, 2010; Zhang *et al.*, 2015) sources. However, it has been suggested that the core human microbial signature is composed of Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria (Buffie and Pamer, 2013). Although, members of the Verrucomicrobia phylum have not been widely reported in WWTP environments, studies have indicated dominance by this phyla in aquatic sources (Brown *et al.*, 2015; Satinsky *et al.*, 2015). Hence, detection of these phyla from both WWTPs and effluent receiving rivers is not surprising. However, shifts in the relative abundances in comparison to the natural population of these phyla in both effluent-receiving rivers (upstream) demonstrate that these WWTPs disrupt the natural bacterial community composition in the respective receiving river sources.

This was also observed from the results in Chapter five where overall, there was a difference in the composition, evenness and diversity indices of the bacterial communities observed across all samples. Here, the bacterial composition for Plant A was higher in the final effluent discharged than in the raw influent received as well as in the effluent-receiving rivers suggesting that the treatment was not efficient in reducing the bacterial load. Analysis of the bacterial communities with more sample sites during the various process and the physicochemical parameters of the wastewater might assist in confirming whether or not the increase in the bacterial richness and diversity in the final effluent discharged is a result of the ineffectiveness of the treatment process. In contrast, Plant B resulted in significantly altered bacterial community composition and diversity downstream of the receiving river. The most striking effect was the absence of bacterial species present in upstream sample of the effluent-

receiving river compared to the downstream sample. In addition, downstream sample was observed to have relatively the same community composition as the wastewater final effluent discharged. The results illustrates the concept of biotic homogenization and suggests that human modifications of the environment are reducing the biological richness that exist in natural ecosystems and consequently only a subset of naturally occurring species are constantly being selected for as a result of the human-altered ecosystems (McKinney 2008; Baiser *et al.* 2012). Effects of this phenomenon has been suggested to result in a more homogenized biosphere with lower diversity on regional and global scales (McKinney and Lockwood, 1999). Therefore, our results suggests that Plant B's effluent may be a driving force for biotic homogenization of bacterial communities in receiving river and may lead to negative implications on the important urban river ecosystem functions.

Results presented in Chapter two highlighted the potential of wastewater effluents in the dissemination of potential pathogenic bacteria into the environment even at low relative abundances. Though the abundance and diversity of the observed potential pathogens declined as a result of the treatment process in both WWTPs, up to 19 genera of pathogenic bacteria were still detected in final treated effluent samples. These included *Legionella*, *Acinetobacter*, *Escherichia*, *Listeria*, *Neisseria*, *Pseudomonas*, *Salmonella*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Vibrio* and *Yersinia*. The presence of *Acinetobacter baumannii* and *Legionella pneumophila* in high frequencies poses potential public health risk as they can easily be disseminated into the river systems. *Acinetobacter baumannii* strains have previously been reported to be present in wastewaters receiving hospital discharges (Ferreira *et al.*, 2011; Zhang *et al.*, 2014) with only a few studies reporting their detection elsewhere in the environment (El-Sayed, 2016; Girlich *et al.*, 2010; Hrenovic *et al.*, 2014). The bacterium has been reported as

one of the most important human pathogens to cause nosocomial pneumonia and bacteraemia in patients admitted in the intensive care unit (Gaynes and Edwards, 2005; Kanafani *et al.*, 2003; Paul *et al.*, 2005; Wisplinghoff *et al.*, 2004) followed by skin, soft tissue, and urinary tract infections (Bergogne-Bérézin and Towner, 1996; Gales *et al.*, 2001) and secondary meningitis (Falagas *et al.*, 2007; Ng *et al.*, 2006) globally. Furthermore, the number of reports of multidrug resistant *A. baumannii* in hospitals has been steadily increasing (Dijkshoorn *et al.*, 2007).

Detection of these potential pathogenic genera in the final treated effluent could be explained by the results presented in Chapter three, where chlorination was found to be ineffective in reducing the pathogenic population due to the presence of several genes involved in the bacterial pathway of glutathione metabolism (Chao *et al.*, 2013), which could offer protective function to these organisms (Chesney *et al.*, 1996). The widely accepted and proposed theory of glutathione metabolism in eukaryotes is by the transfer of these genes from bacteria via the ancestral predecessor of mitochondria during evolution (Fahey *et al.*, 1984). This suggests that the modern relatives of the ancestral predecessor of mitochondria, members of the alpha-proteobacteria class, should house homologous glutathione biosynthesis genes. Moreover, because the nutrient levels present in treated wastewaters prior to the disinfection treatment stage should be at low levels, starvation of bacterial species is expected to stimulate glutathione synthesis and subsequently enhance chlorine resistance (Saby *et al.*, 1999). Additionally, genes involved in glutathione metabolism has been said to be indirectly involved in the regulation of several oxidation resistant systems, such as OxyR, SoxR and SOS systems (Saby *et al.*, 1999), all of which were observed in high abundance in the metagenomes of both WWTPs. Furthermore, genes involved in EPS synthesis, although more related to biofilm production,

could also provide protective functions (Ryu and Beuchat, 2005) and were also observed in the metagenomes of both WWTPs. Therefore, the presence of these genes encoding protective functions could explain the inefficiency of chlorination for inactivation and reduction of the potential pathogens, hence their detection in the final treated effluent.

Annotation, abundance and diversity analysis of genes encoding for antibiotic resistance from the results presented in Chapters two, three and four also revealed changes upon treatment of the raw wastewaters received by both WWTPs. Diversity indices calculations consistently indicated a reduced diversity of ARG types upon wastewater treatment in both WWTPs. This was also observed with PCoA of the ARGs type profiles, with each of the ARG type profiles showing separate clustering between the various wastewater samples. Conversely, slight differences were detected between the upstream and downstream samples along both effluent-receiving rivers. However, no obvious differences between the natural baseline of ARG types in influent samples compared to the final treated effluents and downstream of the effluent-receiving rivers was observed. These observations indicated that the ARGs load from both WWTPs investigated in this study did not have a detectable influence on both the effluent-receiving rivers. This result is in contradiction to the fact that WWTPs represent the main sites through which ARGs are released into the environment (Marti *et al.*, 2014). However, a recent study by Munck and colleagues (2015) suggested that the abundant WWTP resistome is only disseminated, to a very limited extent, to other microbial communities in differing environments. Furthermore, the study suggested that members of the WWTP core resistome rarely take part in gene exchange networks with human pathogens as mobilization seems to be acting as the main barrier preventing the spread of ARGs (Munck *et al.*, 2015). Hence, the findings of this study suggest that the discharge of treated wastewater effluents into receiving

rivers, though introduce ARGs into aquatic sources, did not contribute to increased levels of the natural ARGs. This relatively stable abundance of ARGs could be attributed to the ongoing discharge of inefficiently treated wastewaters by both WWTPs, resulting in an already compounded levels of ARGs within these aquatic sources. Hence, future studies should attempt to corroborate this speculation.

Nonetheless, the proportions of the ARGs identified in this study were diverse and comparable to the results from previous studies investigating activated sludge (Zhang *et al.*, 2011), sewage effluent (Port *et al.*, 2012), plasmids recovered from WWTP (Li et al., 2015), non-hospital medical care facility (Bäumlisberger *et al.*, 2015) and drinking water treatment plants *(Chao et al.*, 2013; Huang *et al.*, 2011; Xi *et al.*, 2009). Specifically, ARGs conferring resistance to the elfamycin and aminocoumarin class of antibiotics was observed to be the most abundant across all wastewater metagenomes from both WWTPs as well as in the respective receiving river bodies. The aminocoumarin family of antibiotics inhibits the beta sub-unit of DNA gyrase which is essential in cell division (Galm *et al.*, 2004). The elfamycin family of antibiotics inhibits protein synthesis of Gram-negative bacteria by interacting with elongation factor Tu (Hall *et al.*, 1989; Vogeley *et al.*, 2001; Wolf *et al.*, 1974). Coincidentally, variants of the elongation factor Tu were the dominant type of elfamycin resistance gene across the wastewater bacterial communities of both WWTPs. Though the elfamycin class of antibiotics are not commonly used in treatment of human diseases due to its narrow spectrum against human pathogens (Miele *et al.*, 1994), it is largely applied as a growth promoting agent in animals owing to its excellent activities (Brötz *et al.*, 2011; Maiese *et al.*, 1989). Though, aminocoumarin is used in the treatment of human diseases, it is also largely applied for treatment of *Stapholococcal* infections, cholera infections and treatment of bovine mastitis in

animal food production (Economou and Gousia, 2015; Marshall and Levy, 2011). Therefore, high abundance of these ARGs across all the metagenomes analysed in this study suggests that a high percentage of the wastewaters received by both WWTPs could be from surrounding animal feed industries with domestic, hospital and other industrial wastewaters contributing to a relatively small percentage. This reason could also partially explain the coinciding high occurrence of genes conferring resistance to tetracycline class of antibiotics as tetracycline is commonly used for livestock purposes in South Africa (Eagar *et al.*, 2012). Furthermore, studies suggests that sources of tetracycline resistance genes could result from clinical (Agersø and Guardabassi, 2005) or animal sources (Agersø and Petersen, 2007) and are common amongst *Acinetobacter* spp. (Hamidian *et al.*, 2016), which is a cause for concern with the increasing threat in hospitals from multidrug-resistant *Acinetobacter* strains (Dijkshoorn *et al.*, 2007). The occurrence of these particular families of antibiotic could also be due to their ineffective biodegradation during the wastewater treatment process and subsequent selections at a low concentrations, persisting in the treated effluents as previously suggested (Gullberg *et al.*, 2011). Overall, the ARGs types detected in the wastewater samples of both WWTPs as well as in the effluent-receiving river bodies were associated with the three common antibiotic resistance mechanisms *viz.* extrusion by efflux pumps, antibiotic inactivation and cellular protection.

Recovery of microbial genomes from the assembly of metagenomic sequence data is of fundamental value to improving our general understanding of microbial ecology and metabolism. This is especially significant in elucidating the functional potential of genomes representing important unculturable microbial lineages that carry out core functions of complex ecosystems such as in aquatic sources. Therefore, Chapter six aimed to recover draft genomes

with as high completeness as possible from the metagenomic datasets in this study. *Polynucleobacter* sp. was recovered with an estimated 68% completeness from the metagenomic datasets. The *Polynucleobacter* genus falls under the Proteobacteria phylum, under the Beta-proteobacteria class and is affiliated with the Burkholderiaceae family. As Proteobacterial species dominated the metagenomes across all samples, it is not surprising that recovery of members within the phylum was observed. This genus, which harbours Gram-negative, non-motile, rod-shaped bacteria was originally established by Heckmann and Schmidt in 1987 (Heckmann and Schmidt, 1987). Organisms from this genus have been reported to be cytoplasmic endosymbiont of the ciliate protist, *Euplotes aediculatus*. To date, there are only four members under this genus *viz. P. necessarius*, *P. cosmopolitanus*, *P. rarus* and *P. acidiphobus*, and by definition, the genus would have been designated as Candidatus (Meincke *et al.*, 2012). Hence, the recovery of PT2 draft genome, identifications and annotations of genes encoded may elucidate the metabolic potential of this organism and contribute to further understandings of its symbiosis with protist. As expected, *Acinetobacter* sp. PT1 was recovered with an estimated 75% completeness, since the genus was observed to be dominant across the metagenomic datasets in this study. *Acinetobacter* is a complex genus which includes Gram-negative, non-motile, aerobic coccobacilli bacteria that are able to inhabit several ecological niches including soil, water, animals and humans (Warskow and Juni, 1972). In addition to the expected genes encoding for catabolism, biosynthesis, and stress response, the PT1 draft genome was annotated with genes encoding prophages, transposable elements and plasmids, with genes under virulence, disease and defence category also detected. Furthermore, several resistance genes encoding for β-lactamases ($bla_{OXA-211}$, $bla_{OXA-212}$ and $bla_{OXA-309}$), macrolide resistance (*mphE* and *msrE*), and tetracycline resistance determinant (*tet39*), which have been reported to be found in several *Acinetobacter* species (Figueiredo *et al.*, 2012) were also identified in the PT1 draft genome. Identification of the draft genome with

the GenBank database showed best hits for *Acinetobacter* spp. Furthermore, resolving to the species level placed PT1 to be closely related to well-studied *A. johnsonii*, suggesting PT1 to be members of the *A. johnsonii* species. Alarmingly, *A. baumannii* have been associated with nosocomial infections, with increasing number of multi-drug resistant strains, resulting in high rates of mortality and long periods of hospitalization (Dijkshoorn *et al.*, 2007; Falagas *et al.*, 2006). Hence, the recovery of PT1 draft genome and annotations of genes identified would contribute to further understandings of this genus and provide possible targets for treatment of its related species.

7.2 Potential for future developments

The omics field is rapidly expanding due to the exponentially decreasing cost of sequencing and widespread availability of sequencers and mass spectrometers coupled with the seemingly unlimited breadth of applications. However, generating, processing, analysing, interpreting data and inferring biological meaning can still take months to years and requires substantial technical and general expertise in large teams due to the swift evolving nature of these techniques. In this study, only one of the five omic based approaches (meta-genomics, -transcriptomics, -proteomics, -bolomics and -phenomics) was utilised. Although the wealth of information generated from metagenomics is quite substantial and proven to be highly informative, the DNA sequence-based method still has limitations. For example, taxonomic resolution can only go up to species level at best. However, it is known that many important phenomena such as the acquisition of antibiotic resistance genes occurs at the strain level.

Therefore, in order to understand the evolutionary and epidemiological relationships among bacterial pathogens, whole genome analysis has been previously used to identify and attribute the outbreak sources for many bacterial pathogenic strains, including *Escherichia coli* O104 (Rasko *et al.*, 2011a), *Vibrio cholerae* (Hendriksen *et al.*, 2011), *Klebsiella* spp. (Snitkin *et al.*, 2012) methicillin resistant *Staphylococcus aureus* (MRSA) (Price et al., 2012) and *Bacillus anthracis* (Rasko *et al.*, 2011b). However, an approach with the power to resolve complex microbial communities at a strain level to determine genomic variations has only recently been proposed by Eren *et al.* (2013). This strategy termed "oligotyping" is a supervised computational method that allows for investigation in the diversity of closely related but distinct bacteria through canonical approaches in the final operational taxonomic units established from amplicon sequencing of the conserved 16S ribosomal RNA gene data (Eren *et al.*, 2013). This approach has been applied in studies investigating the structure of the vaginal microbiome and was able to determine 46 unique oligotypes within the *Gardnerella vaginalis* species (Eren *et al.*, 2011). Another approach, demonstrated by the Human Microbiome Project, showed that mapping of shotgun sequencing reads from tongue samples to genomes of *Streptococcus mitis* was able to reveal the absence or presence of genomic islands within strains of the same species identified from individuals enrolled in the project. These genomic islands were shown to contain multiple and functionally coherent genes that were gained and lost together, suggesting a mechanism for individual-specific and body site-specific functional specialisation of *S. mitis* (Consortium, 2012). In the same project, where greater shotgun sequencing depths of the microbiome was analysed, determination of single-nucleotide polymorphisms (SNPs) were possible from human stool samples and used for comparisons to high quality known reference genomes. This analysis revealed that subject-specific SNPs variations tend to remain stable for up to one year and was comparatively more conserved than the overall species abundance. Furthermore, ranking of species and genes in the gut by the

degree of SNP between the individuals revealed that antibiotic resistance genes were amongst the most variable whilst the housekeeping genes were the most conserved (Schloissnig *et al.*, 2013). Therefore, further analysis of the metagenomic datasets sequenced in this study should move towards this direction in order to further elucidate the pathogenic bacterial populations in the metagenomes.

Conversely, the importance of integrated multi-omic studies have recently gained the spotlight leading to some momentum and future investigations should move towards this direction. For example, a study aimed at targeting the functional role of Oceanospirillales species as well as other active members of the indigenous microbial community following the 2010 Deepwater Horizon oil spill which has potential in aiding bioremediation of the spillage. This was achieved by utilising the combination of meta-genomics, -transcriptomics as well as single-cell genomics. Metagenomic and metatranscriptomic sequencing revealed that genes for motility, chemotaxis and aliphatic hydrocarbon degradation were significantly enriched and expressed in the hydrocarbon plume samples compared with uncontaminated seawater collected from plume depth. In contrast, although genes coding for degradation of recalcitrant compounds, such as benzene, toluene, ethylbenzene, total xylenes and polycyclic aromatic hydrocarbons, were identified from the metagenomes, they were expressed at low levels or not at all based on analysis of the metatranscriptomes. Isolation and genome sequencing of two Oceanospirillales single cells revealed genes coding for n-alkane and cycloalkane degradation with near-complete pathway for cyclohexane oxidation and supported the detection in both metagenome and metatranscriptome data analysis. Furthermore, the draft genomes also included genes for chemotaxis, motility and nutrient acquisition strategies which were also determined in the metagenomes and metatranscriptomes (Mason *et al.*, 2012). Therefore, integrated omics allows

for simultaneous enhanced understanding of not only the community structure but also the *in vivo* function and dynamics of the community studied. Hence, future studies should entail a multi-omic approach utilizing meta-genomic, -transcriptomic as well as single-cell genomics. As demonstrated in this study, metagenomics would allow for elucidation of the complex bacterial communities present, the small pathogenic population, metabolic potential of the communities as well as the antibiotic resistome, whilst metatranscriptomics would validate the functional genes determined from metagenomics and indicate which of the enriched genes are expressed resulting in the protective functions and resistance to treatment described in Chapter three. Furthermore, metatranscriptomics and single-cell genomics would reveal the resistance genes expressed and mechanisms of horizontal gene transfer as well as isolation and subsequent reconstruction of complete genomes of the small pathogenic population, respectively. Combined, these results will allow for complete understanding of pathogenic population within the particular WWTP of concern and provide knowledge which may be useful when designing effective wastewater treatment strategies, thereby reducing the risks associated with improperly treated wastewaters discharged into the receiving rivers. Hence, it has been suggested that simultaneous generation of meta-omics data is crucial to fully understand the functional capacity of microbial communities (Muller *et al.*, 2014). Furthermore, systematically obtained temporal- and spacial-resolved omic datasets will allow for establishment of structure to function relationships by identifying key members and its functions. Such knowledge also offers a starting point in potentially novel biological-functional discoveries and is attractive for biotechnological applications (Narayanasamy *et al.*, 2015).

Next-generation sequencing platforms utilising single molecule real-time technology such as the SMRT from Pacific Biosciences and MinION from Oxford Nanopore Technologies are

said to be capable of producing sequences with read lengths greater than 1000 bp and produce reads of up to 99.99% consensus accuracy, overall cost-effectiveness and small pocket-sized mobility, surpassing some of the limitations from the Illumina and Roche platforms. This coupled with real-time data acquisition makes both platforms attractive for future environmental monitoring applications. Despite these advantages, these platforms have not been widely adopted to date in surveys of highly diverse bacterial communities due to initial studies revealing high rates of randomly distributed sequencing errors which could lead to artificially inflated community diversity (Schloss *et al.*, 2016). Nevertheless, recent studies employing the SMRT platform with improved sequencer hardware and chemistry yielded improved accuracy with read lengths greater than 1400 bp (Marshall *et al.*, 2012; Mosher *et al.*, 2014). Though riddled with a high sequence error rate of approximately 30%, sequencing results produced by laboratories which participated in the MinION Access Program utilising the platform in a plethora of different environmental and human samples showed that MinION could generate single read lengths of up to 5500 bp in a single run (Ashton *et al.*, 2015; Madoui *et al.*, 2015; Mikheyev and Tin, 2014). Therefore, future investigations in determining the bacterial pathogens should utilised these emerging NGS platforms in combinations with the established Illumina NGS platforms. These studies would allow for technical limitations to be overcome and will ultimately open up new avenues of monitoring water quality and ensuring sustainable wastewater treatment.

7.3 Concluding remarks

Microbial water quality assessment through NGS-based molecular detection of community compositions, potential pathogens as well as genes encoding for antibiotic resistance or important biogeochemical processes, have the potential to be translated into actionable data for water quality managers. The overall findings of this study demonstrates the usefulness of NGS in initial evaluations of microbial contaminants in wastewater effluents during water quality assessments as well as the possible impacts on the receiving river bodies. This study indicates that urban WWTPs are indeed potential hotspots for antibiotic resistant bacteria and antibiotic resistance genes. Furthermore, inefficiency of the treatment process could contribute towards the dissemination of potentially pathogenic and multidrug resistant bacteria into the environment. Moreover, genes encoding for resistance to major classes of antibiotics detected in treated effluents of these urban WWTPs could contribute to the rise in global antibiotic resistance. Hence, constant monitoring is crucial to safeguard freshwater resources as well as the public human health. However, crucial relationships between the occurrence, detection and quantification of nucleic acids in the environment through NGS approaches and potential impacts on human or environmental health still needs to be further investigated before profiles based on the distribution of microbial genes are relied upon to replace currently used indicator organisms and direct pathogen detections. Current frameworks for water quality assessment heavily rely upon quantifications of these indicator organisms with decades of literature validating its approach coupled with epidemiological data. Hence, future research needs to investigate whether the utilization of NGS for water quality assessment can improve the existing frameworks.

7.4 References

Agersø, Y., Guardabassi, L., 2005. Identification of Tet 39, a novel class of tetracycline resistance determinant in *Acinetobacter* spp. of environmental and clinical origin. J. Antimicrob. *Chemother*. 55, 566–569.

Agersø, Y., Petersen, A., 2007. The tetracycline resistance determinant Tet 39 and the sulphonamide resistance gene *sulII* are common among resistant *Acinetobacter* spp. isolated from integrated fish farms in Thailand. J. Antimicrob. Chemother. 59, 23–27.

Andreote, F.D., Jiménez, D.J., Chaves, D., Dias, A.C.F., Luvizotto, D.M., Dini-Andreote, F., Fasanella, C.C., Lopez, M.V., Baena, S., Taketani, R.G., de Melo, I.S., 2012. The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One* 7, e38600.

Ashton, P.M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J., O'Grady, J., 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* 33, 296–300.

Baiser, B., Olden, J.D., Record, S., Lockwood, J.L., McKinney, M.L., 2012. Pattern and process of biotic homogenization in the New Pangaea. *Proc. Biol. Sci.* 279, 4772–4777.

Bergogne-Bérézin, E., Towner, K.J., 1996. *Acinetobacter* spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features. *Clin. Microbiol. Rev.* 9, 148–165.

Brown, B.L., LePrell, R. V, Franklin, R.B., Rivera, M.C., Cabral, F.M., Eaves, H.L., Gardiakos, V., Keegan, K.P., King, T.L., 2015. Metagenomic analysis of planktonic microbial consortia from a non-tidal urban-impacted segment of James River. *Stand. Genomic Sci.* 10, 65.

Brötz, E., Kulik, A., Vikineswary, S., Lim, C.-T., Tan, G.Y.A., Zinecker, H., Imhoff, J.F., Paululat, T., Fiedler, H.-P., 2011. Phenelfamycins G and H, new elfamycin-type antibiotics produced by *Streptomyces albospinus* Acta 3619. J. Antibiot. (Tokyo). 64, 257–66. doi:10.1038/ja.2010.170

Buffie, C.G., Pamer, E.G., 2013. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* 13, 790–801.

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X., Wu, W., Zhang, T., 2013. Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. *Sci. Rep.* 3, 3550.

Chesney, J.A., Eaton, J.W., Mahoney, J.R., 1996. Bacterial glutathione: a sacrificial defense against chlorine compounds. *J. Bacteriol.* 178, 2131–5.

Consortium, H.M.P., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.

Corcoran, E., Nellemann, C., Baker, E., Bos, R., Osborn, D., Savelli, H., 2010. Sick Water? The central role of Response, management in sustainable development. A Rapid Programme, UNEP/UN-HABITAT, Birkeland Trykkeri AS, Norway.

Da Silva, M.L.B., Cantão, M.E., Mezzari, M.P., Ma, J., Nossa, C.W., 2015. Assessment of bacterial and archaeal community structure in Swine wastewater treatment processes. *Microb. Ecol.* 70, 77–87.

Department of Environmental Affairs, 2016. 2nd South Africa Environment Outlook - A report on the state of the environment. Pretoria, South Africa.

Dijkshoorn, L., Nemec, A., Seifert, H., 2007. An increasing threat in hospitals: Multidrug-resistant *Acinetobacter baumannii*. *Nat. Rev. Microbiol.* 5, 939–951.

Eagar, H., Swan, G., van Vuuren, M., 2012. A survey of antimicrobial usage in animals in South Africa with specific reference to food animals. *J. S. Afr. Vet. Assoc.* 83, 16.

Economou, V., Gousia, P., 2015. Agriculture and food animals as a source of antimicrobial-resistant bacteria. *Infect. Drug Resist.* 8, 49–61.

El-Sayed, M.H., 2016. Multiple heavy metal and antibiotic resistance of *Acinetobacter baumannii* Strain HAF – 13 isolated from industrial effluents. *Am. J. Microbiol. Res.* 4, 26–36.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., Sogin, M.L., 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* 4, 1111–1119.

Eren, A.M., Zozaya, M., Taylor, C.M., Dowd, S.E., Martin, D.H., Ferris, M.J., 2011. Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS One* 6, e26732.

Ettema, T.J.G., Andersson, S.G.E., 2009. The alpha-proteobacteria: the Darwin finches of the bacterial world. *Biol. Lett.* 5, 429–432.

Fahey, R.C., Newton, G.L., Arrick, B., Overdank-Bogart, T., Aley, S.B., 1984. *Entamoeba histolytica*: A eukaryote without glutathione metabolism. *Science* 224, 70–2.

Falagas, M.E., Bliziotis, I.A., Siempos, I.I., 2006. Attributable mortality of *Acinetobacter baumannii* infections in critically ill patients: A systematic review of matched cohort and case-control studies. *Crit. Care* 10, R48.

Falagas, M.E., Bliziotis, I.A., Tam, V.H., 2007. Intraventricular or intrathecal use of polymyxins in patients with Gram-negative meningitis: A systematic review of the available evidence. *Int. J. Antimicrob. Agents* 29, 9–25.

Ferreira, A.E., Marchetti, D.P., De Oliveira, L.M., Gusatti, C.S., Fuentefria, D.B., Corção, G., 2011. Presence of OXA-23-producing isolates of *Acinetobacter baumannii* in wastewater from hospitals in southern Brazil. *Microb. Drug Resist.* 17, 221–227.

Figueiredo, S., Bonnin, R.A., Poirel, L., Duranteau, J., Nordmann, P., 2012. Identification of the naturally occurring genes encoding carbapenem-hydrolysing oxacillinases from *Acinetobacter haemolyticus*, *Acinetobacter johnsonii*, and *Acinetobacter calcoaceticus*. *Clin. Microbiol. Infect.* 18, 907–913.

Foong, C.P., Wong Vui Ling, C.M., González, M., 2010. Metagenomic analyses of the dominant bacterial community in the Fildes Peninsula, King George Island (South Shetland Islands). *Polar Sci.* 4, 263–273.

Gales, A.C., Jones, R.N., Forward, K.R., Liñares, J., Sader, H.S., Verhoef, J., 2001. Emerging importance of multidrug-resistant *Acinetobacter* species and *Stenotrophomonas maltophilia* as pathogens in seriously ill patients: Geographic patterns, epidemiological features, and trends in the SENTRY Antimicrobial Surveillance Program (1997-1999. *Clin. Infect. Dis.* 32, S104–S113.

Galm, U., Heller, S., Shapiro, S., Page, M., Li, S., Heide, L., 2004. Antimicrobial and DNA gyrase-inhibitory activities of novel clorobiocin derivatives produced by mutasynthesis. *Antimicrob. Agents Chemother.* 48, 1307–1312.

Gaynes, R., Edwards, J.R., 2005. Overview of nosocomial infections caused by Gram-negative bacilli. *Clin. Infect. Dis.* 41, 848–854.

Ghai, R., Rodriguez-Valera, F., McMahon, K.D., Toyama, D., Rinke, R., Cristina Souza de Oliveira, T., Wagner Garcia, J., Pellon de Miranda, F., Henrique-Silva, F., 2011. Metagenomics of the water column in the pristine upper course of the Amazon river. *PLoS One* 6, e23785.

Girlich, D., Poirel, L., Nordmann, P., 2010. First isolation of the *blaOXA-23* carbapenemase gene from an environmental *Acinetobacter baumannii* isolate. *Antimicrob. Agents Chemother*. 54, 578–579.

Gullberg, E., Cao, S., Berg, O.G., Ilbäck, C., Sandegren, L., Hughes, D., Andersson, D.I., 2011. Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog.* 7, e1002158.

Guo, J., Peng, Y., Ni, B., Han, X., Fan, L., Yuan, Z., 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. *Microb. Cell Fact.* 14, 33.

Hall, C.C., Watkins, J.D., Georgopapadakou, N.H., 1989. Effects of elfamycins on elongation factor Tu from *Escherichia coli* and *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 33, 322–325.

Hamidian, M., Holt, K.E., Pickard, D., Hall, R.M., 2016. A small *Acinetobacter* plasmid carrying the *tet39* tetracycline resistance determinant. *J. Antimicrob. Chemother.* 71, 269–271.

Heckmann, K., Schmidt, H.J., 1987. *Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. *Int. J. Syst. Bacteriol.* 37, 456–457.

Hendriksen, R.S., Price, L.B., Schupp, J.M., Gillece, J.D., Kaas, R.S., Engelthaler, D.M., Bortolaia, V., Pearson, T., Waters, A.E., Upadhyay, B.P., Shrestha, S.D., Adhikari, S., Shakya, G., Keim, P.S., Aarestrup, F.M., 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the origin of the Haitian outbreak. *MBio* 2, e00157–11.

Hrenovic, J., Durn, G., Goic-Barisic, I., Kovacic, A., 2014. Occurrence of an environmental *Acinetobacter baumannii* strain similar to a clinical isolate in paleosol from Croatia. *Appl. Environ. Microbiol.* 80, 2860–2866.

Hu, M., Wang, X., Wen, X., Xia, Y., 2012. Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresour. Technol.* 117, 72–79.

Huang, J., Hu, H., Tang, F., Li, Y., Lu, S., Lu, Y., 2011. Inactivation and reactivation of antibiotic-resistant bacteria by chlorination in secondary effluents of a municipal wastewater treatment plant. *Water Res.* 45, 2775–2781.

Jordaan, K., Bezuidenhout, C., 2013. The impact of physico-chemical water quality parameters on bacterial diversity in the Vaal River, South Africa. *Water S.A.* 39, 385–396.

Kanafani, Z.A., Kara, L., Hayek, S., Kanj, S.S., 2003. Ventilator-associated pneumonia at a tertiary-care center in a developing country: Incidence, microbiology, and susceptibility patterns of isolated microorganisms. *Infect. Control Hosp. Epidemiol.* 24, 864–869.

Kolmakova, O. V, Gladyshev, M.I., Rozanov, A.S., Peltek, S.E., Trusova, M.Y., 2014. Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. *FEMS Microbiol. Ecol.* 89, 442–450.

Li, A., Li, L., Zhang, T., 2015. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. *Front. Microbiol.* 6, 1025.

Madoui, M., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P., Aury, J.-M., 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* 16, 327.

Maiese, W.M., Lechevalier, M.P., Lechevalier, H.A., Korshalla, J., Goodman, J., Wildey, M.J., Kuck, N., Conner, S.D., Greenstein, M., 1989. LL-E19020 alpha and beta, animal growth promoting antibiotics: taxonomy, fermentation and biological activity. J. Antibiot. 42, 1489–1493.

Marshall, B.M., Levy, S.B., 2011. Food animals and antimicrobials: Impacts on human health. *Clin. Microbiol. Rev.* 24, 718–733.

Marshall, C.W., Ross, D.E., Fichot, E.B., Norman, R.S., May, H.D., 2012. Electrosynthesis of commodity chemicals by an autotrophic microbial community. *Appl. Environ. Microbiol.* 78, 8412–8420.

Marti, E., Variatza, E., Balcazar, J.L., 2014. The role of aquatic ecosystems as reservoirs of antibiotic resistance. *Trends Microbiol.* 22, 36–41.

Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S.G., Dubinsky, E.A., Fortney, J.L., Han, J., Holman, H.-Y.N., Hultman, J., Lamendella, R., Mackelprang, R., Malfatti, S., Tom, L.M., Tringe, S.G., Woyke, T., Zhou, J., Rubin, E.M., Jansson, J.K., 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* 6, 1715–1727.

McKinney, M., Lockwood, J., 1999. Biotic homogenization: A few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* 14, 450–453.

McKinney, M.L., 2008. Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosyst.* 11, 161–176.

Meincke, L., Copeland, A., Lapidus, A., Lucas, S., Berry, K.W., Del Rio, T.G., Hammon, N., Dalin, E., Tice, H., Pitluck, S., Richardson, P., Bruce, D., Goodwin, L., Han, C., Tapia, R., Detter, J.C., Schmutz, J., Brettin, T., Larimer, F., Land, M., Hauser, L., Kyrpides, N.C., Ivanova, N., Göker, M., Woyke, T., Wu, Q.L., Pöckl, M., Hahn, M.W., Klenk, H., 2012. Complete genome sequence of *Polynucleobacter necessarius* subsp. *asymbioticus* type strain (QLW-P1DMWA-1^{T}). *Stand. Genomic Sci.* 6, 74–83.

Miele, A., Goldstein, B.P., Bandera, M., Jarvis, C., Resconi, A., Williams, R.J., 1994. Differential susceptibilities of enterococcal species to elfamycin antibiotics. *J. Clin. Microbiol.* 32, 2016–2018.

Mikheyev, A.S., Tin, M.M.Y., 2014. A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* 14, 1097–1102.

Miura, Y., Hiraiwa, M.N., Ito, T., Itonaga, T., Watanabe, Y., Okabe, S., 2007. Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance. *Water Res.* 41, 627–637.

Mosher, J.J., Bowman, B., Bernberg, E.L., Shevchenko, O., Kan, J., Korlach, J., Kaplan, L.A., 2014. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *J. Microbiol. Methods* 104, 59–60.

Muller, E.E.L., Sheik, A.R., Wilmes, P., 2014. Lipid-based biofuel production from wastewater. *Curr. Opin. Biotechnol.* 30, 9–16.

Munck, C., Albertsen, M., Telke, A., Ellabaan, M., Nielsen, P.H., Sommer, M.O.A., 2015. Limited dissemination of the wastewater treatment plant core resistome. *Nat. Commun.* 6, 8452.

Narayanasamy, S., Muller, E.E.L., Sheik, A.R., Wilmes, P., 2015. Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microb. Biotechnol*. 8, 363–8.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L., Knight, R., 2011. Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* 13, 135–144.

Ng, J., Gosbell, I.B., Kelly, J.A., Boyle, M.J., Ferguson, J.K., 2006. Cure of multiresistant *Acinetobacter baumannii* central nervous system infections with intraventricular or intrathecal colistin: Case series and literature review. *J. Antimicrob. Chemother.* 58, 1078–1081.

Paul, M., Weinberger, M., Siegman-Igra, Y., Lazarovitch, T., Ostfeld, I., Boldur, I., Samra, Z., Shula, H., Carmeli, Y., Rubinovitch, B., Pitlik, S., 2005. *Acinetobacter baumannii*: Emergence and spread in Israeli hospitals 1997-2002. *J. Hosp. Infect.* 60, 256–260.

Port, J.A., Wallace, J.C., Griffith, W.C., Faustman, E.M., 2012. Metagenomic profiling of microbial composition and antibiotic resistance determinants in Puget Sound. *PLoS One* 7, e48000.

Price, L.B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P.S., Pearson, T., Waters, A.E., Foster, J.T., Schupp, J., Gillece, J., Driebe, E., Liu, C.M., Springer, B., Zdovc, I., Battisti, A., Franco, A., Zmudzki, J., Schwarz, S., Butaye, P., Jouy, E., Pomba, C., Porrero, M.C., Ruimy, R., Smith, T.C., Robinson, D.A., Weese, J.S., Arriola, C.S., Yu, F., Laurent, F., Keim, P., Skov, R., Aarestrup, F.M., 2012. *Staphylococcus aureus* CC398: Host adaptation and emergence of methicillin resistance in livestock. *MBio* 3, e00305-11.

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E.E., Sebra, R., Chin, C., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A.O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J.C., Steyert, S.R., Frimodt-Møller, J., Struve, C., Petersen, A.M., Krogfelt, K.A., Nataro, J.P., Schadt, E.E., Waldor, M.K., 2011a. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J. Med.* 365, 709–17.

Rasko, D.A., Worsham, P.L., Abshire, T.G., Stanley, S.T., Bannan, J.D., Wilson, M.R., Langham, R.J., Decker, R.S., Jiang, L., Read, T.D., Phillippy, A.M., Salzberg, S.L., Pop, M., Van Ert, M.N., Kenefic, L.J., Keim, P.S., Fraser-Liggett, C.M., Ravel, J., 2011b. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5027–32.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1, 283–290.

Ryu, J., Beuchat, L.R., 2005. Biofilm formation by *Escherichia coli* O157:H7 on stainless steel: Effect of exopolysaccharide and curli production on its resistance to chlorine. *Appl. Environ. Microbiol.* 71, 247–54.

Saby, S., Leroy, P., Block, J.C., 1999. *Escherichia coli* resistance to chlorine and glutathione synthesis in response to oxygenation and starvation. *Appl. Environ. Microbiol.* 65, 5600–3.

Sánchez-Andrea, I., Rodríguez, N., Amils, R., Sanz, J.L., 2011. Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. *Appl. Environ. Microbiol.* 77, 6085–6093.

Satinsky, B.M., Fortunato, C.S., Doherty, M., Smith, C.B., Sharma, S., Ward, N.D., Krusche, A. V, Yager, P.L., Richey, J.E., Moran, M.A., Crump, B.C., 2015. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome* 3, 39.

Sauvain, L., Bueche, M., Junier, T., Masson, M., Wunderlin, T., Kohler-Milleret, R., Gascon Diez, E., Loizeau, J., Tercier-Waeber, M., Junier, P., 2013. Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. *Aquat. Sci.* 76, 33–46.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D.R., Kultima, J.R., Martin, J., Kota, K., Sunyaev, S.R., Weinstock, G.M., Bork, P., 2013. Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50.

Schloss, P.D., Jenior, M.L., Koumpouras, C.C., Westcott, S.L., Highlander, S.K., 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4, e1869.

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl. Environ. Microbiol.* 79, 2906–2913.

Shi, P., Jia, S., Zhang, X., Zhang, T., Cheng, S., Li, A., 2013. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water Res.* 47, 111–120.

Silva, C.C., Jesus, E.C., Torres, A.P.R., Sousa, M.P., Santiago, V.M.J., Oliveira, V.M., 2010. Investigation of bacterial diversity in membrane bioreactor and conventional activated sludge processes from petroleum refineries using phylogenetic and statistical approaches. *J. Microbiol. Biotechnol.* 20, 447–459.

Snaidr, J., Amann, R., Huber, I., Ludwig, W., Schleifer, K.H., 1997. Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. *Appl. Environ. Microbiol.* 63, 2884–2896.

Snitkin, E.S., Zelazny, A.M., Thomas, P.J., Stock, F., NISC Comparative Sequencing Program Group, Henderson, D.K., Palmore, T.N., Segre, J.A., 2012. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* 4, 148ra116.

Vogeley, L., Palm, G.J., Mesters, J.R., Hilgenfeld, R., 2001. Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox. *J. Biol. Chem.* 276, 17149–17155.

Wang, Z., Zhang, X., Huang, K., Miao, Y., Shi, P., Liu, B., Long, C., Li, A., 2013. Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. *PLoS One* 8, e76079.

Warskow, A.L., Juni, E., 1972. Nutritional requirements of *Acinetobacter* strains isolated from soil, water, and sewage. *J. Bacteriol.* 112, 1014–1016.

Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P., Edmond, M.B., 2004. Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.* 39, 309–317.

Wolf, H., Chinali, G., Parmeggiani, A., 1974. Kirromycin, an inhibitor of protein biosynthesis that acts on elongation factor Tu. *Proc. Natl. Acad. Sci. U.S.A.* 71, 4910–4914.

World Economic Forum, 2015. Global Risks 2015, 10th edn. Geneva, Switzerland.

Wu, H., Zhao, H., Wen, C., Guo, Y., Guo, J., Xu, M., Li, X., 2012. A comparative study of bacterial community structures in the sediments from brominated flame retardants contaminated river and non-contaminated reservoir. *African J. Microbiol. Res.* 6, 3248–3260.

Xi, C., Zhang, Y., Marrs, C.F., Ye, W., Simon, C., Foxman, B., Nriagu, J., 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. *Appl. Environ. Microbiol.* 75, 5714–5718.

Xia, S., Duan, L., Song, Y., Li, J., Piceno, Y.M., Andersen, G.L., Alvarez-Cohen, L., Moreno-Andrade, I., Huang, C., Hermanowicz, S.W., 2010. Bacterial community structure in geographically distributed biological wastewater treatment reactors. *Environ. Sci. Technol.* 44, 7391–7396.

Yu, K., Zhang, T., 2012. Metagenomic and Metatranscriptomic Analysis of microbial community structure and gene expression of activated sludge. *PLoS One* 7, e38183.

Zhang, C., Qiu, S., Wang, Y., Qi, L., Hao, R., Liu, X., Shi, Y., Hu, X., An, D., Li, Z., Li, P., Wang, L., Cui, J., Wang, P., Huang, L., Klena, J.D., Song, H., 2014. Higher isolation of NDM-1 producing *Acinetobacter baumannii* from the sewage of the hospitals in Beijing. *PLoS One* 8, e64857.

Zhang, T., Zhang, X., Ye, L., 2011. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS One* 6, e26041.

Zhang, Y., Cong, J., Lu, H., Li, G., Xue, Y., Deng, Y., Li, H., Zhou, J., Li, D., 2015. Soil bacterial diversity patterns and drivers along an elevational gradient on Shennongjia Mountain, China. *Microb. Biotechnol.* 8, 739–746.