EFFECT OF POLLUTION GRADIENT ON SEDIMENT BACTERIAL COMPOSITION AND POTENTIAL PATHOGENS IN URBANIZATION-IMPACTED STREAMS DRAINING INTO LAKE VICTORIA

ODHIAMBO KENNEDY ACHIENG'

S151/4151/2017

A THESIS SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF A DEGREE OF MASTER OF SCIENCE DEGREE IN MICROBIOLOGY OF JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

NOVEMBER 2022

DECLARATION AND RECOMMENDATION

Declaration

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

Kennedy Odhiambo Achieng'
S151/4151/2017

Signature: ………………………… Date: …………………………….………...

Recommendation

This research thesis has been submitted for examination with our approval as university supervisors.

Dr. Henry Joseph Oduor Ogola, PhD

Department of Environmental Science,

University of South Africa.

Signature: ……………………… Date: …………………………….………...

Dr. Benson Onyango, PhD

Department of Biological and Physical Sciences,

Jaramogi Oginga Odinga University of Science and Technology.

Signature: ………………………. Date: …………………………….………...

ACKNOWLEDGEMENT

DEDICATION

It is with genuine gratitude and warm regard that I dedicate this thesis to Almighty God, my creator, and source of wisdom and knowledge. He has been the source of my strength throughout this study and has made this work successful by bringing the right people on my side. To begin, I'd like to thank my wonderful teacher and mentor, Dr. Henry JO Ogola, for providing me with hope and wonderful mentorship throughout this work. My family, who have been a great help to me by encouraging me all the way, made sure that I gave it all it took to complete this research. I also dedicate this work to my brother, Mike, who stands by me when things look bleak. Thank you. My love for you all can never be quantified. God bless you.

ABSTRACT

Despite urban rivers/streams draining into Lake Victoria suffering from urbanization-linked anthropogenic pollution, little is known about their microbiome diversity and structure, or how they respond to intensive anthropogenic inputs. This study conducted a comprehensive analysis of the spatial bacterial community distribution in the sediments of Kisat and Auji streams, that flows through Kisumu City into Lake Victoria's Winam Gulf. Specifically, the study: i) used 16S rRNA gene-based Illumina MiSeq sequencing to determine the diversity and abundance of sediment bacterial communities along the stream catchment impacted by varying levels of urbanisation; ii) determined the presence of potential pathogens and the predicted functional profiles of the sediment bacterial communities in the river to establish their role in the ecosystem, and (iii) identified the key environmental factors (nutritional factors and heavy metals) influencing compositional variations in these communities. The study adopted a stratified purposive sampling, where 22 sediment samples were randomly collected from Lower, Mid and Upper catchment of Auji and Kisat streams, stratified as highly, moderate and non-urbanized zones, respectively, based on land use patterns. Results showed that polluted mid and lower catchment zones stream sediments were highly enriched ($p < 0.05$) with *Actinobacteria and Proteobacteria,* and potential pathogen groups such *Corynebacterium*, *Staphylococcus*, *Cutibacterium*, *Turicella*, *Acinetobacter* and *Micrococcus,* including enterics such as *Faecalibacterium*, *Escherichia*, *Klebsiella*, *Enterococcus*, *Prevotella*, *Legionella*, *Vibrio* and *Salmonella.* Further, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) analysis revealed enrichment of genes associated with carbon and nitrogen metabolism and disease pathogenesis and virulence in the lower and mid catchment zones stream sediment. Physicochemical analyses also showed that the highly urbanized mid and lower stream catchment zones had significantly higher ($p < 0.05$) total organic carbon (TOC), total nitrogen (TN), and total phosphorous (TP) content, including severely pollution with toxic heavy metals such as lead (Pb), cadmium (Cd), and copper (Cu) than the less urbanized upper catchment zone. Multivariate analysis suggested that TOC, Pb, Cd, TN, pH and Cr were the significant drivers ($p < 0.01$) of spatial variation in community structure, with Pb, TOC and Cd content being most influential sediment properties ($p < 0.01$). Overall, these results suggest urban pollution significantly affects the stream sediment microbiome and that the current waste management in Kisumu City is insufficient for the protection of public health and aquatic ecosystems. Therefore, proactive and sustainable urban waste management strategies are needed as the city undergoes rapid urbanization.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF PLATES AND FIGURES

ABBREVIATIONS AND ACRONYMS

ANOSIM   Analysis of Similarity

BOD   Biological Oxygen Demand

COD   Chemical Oxygen Demand

DOD   Dissolved Oxygen Demand

FDR   False Discovery Rate

HBPD   Human Bacterial Pathogens Database

KEBS   Kenya Bureau of Standards

KEGG   Kyoto Encyclopaedia Genes and Genomes

LDA   Linear Discriminant Analysis

LEfSe   Linear Discriminant Analysis Effect Size

MST   Microbial Source Tracking

NEMA   National Environment Management Authority

NCBI-BLAST National Center for Biotechnology Information-basic local alignment search tool

NMDS   Non-metric multidimensional scaling

NTU   Nephelometric Turbidity Unit

OTU   Operational Taxonomic Unit

PAHs   Polycyclic aromatic hydrocarbons

PCBs   Polychlorinated Biphenyls

POP   Persistent Organic Pollutants

PERMANOVA Permutational Multivariate Analysis of Variance

PICRUSt   Phylogenetic Investigation of Communities by Reconstruction of Unobserved States

RDA   Redundancy Analysis

TN   Total Nitrogen

TOC   Total Organic Carbon

UNICEF   United Nations International Children Emergency Fund

USDA NRCS United States Department of Agriculture Natural Resource Conservation Services

WHO   World Health Organization

WASREB   Water Services Regulatory Board

WWTP   Waste Water Treatment Plant

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Globally, the majority of human settlement is increasingly moving toward urbanization, with 70% of the world's population anticipated to live in cities by 2050 (Angel et al., 2012). However, the growing city populations, the process of urbanization and its associated land-use change is placing increased pressures on urban infrastructure, utilities and services that may substantially alter urban water, an important interface between human activity and the natural environment (McLellan et al., 2015; Numberger et al., 2021). Rivers and streams flowing through cities are continuously under threat from multiple environmental stressors, as they are important sinks for waste, sewage, and storm water runoff. The geomorphology, physicochemical properties, eukaryotic and microbial community structure, function, and health of rivers and streams can all be altered by this pollution, as well as their structure, composition, and function (Schwarzenbach *et al*., 2006). The problem of urban pollution of rivers and streams is high in cities within developing countries, which lack effective waste management strategies, and has become a major public concern in most urban metropolis (UNESCO, 2003).

In riverine ecosystem, sediments serve as primary sites for the accumulation and attachment of microorganisms and chemical pollutants. As a consequence, sediment microbial communities are more tightly related to environmental conditions within a river/stream since these habitats are more stable and bacterial residency time is greater (Beattie et al., 2020). Sediment microbiome composition and biodiversity are highly sensitive to anthropogenic disturbances (McLellan *et al*., 2015). As a result, altering the sediment microbiome can reveal important information about the impact of urban pollution on stream ecosystems, the presence of health hazards, or even significant clues about short- and long-term ecosystem change (Cui et al., 2019; Fisher et al., 2015; McLellan et al., 2015; Numberger et al., 2021). In 8 rivers within the Parkers Creek watershed in Michigan, USA, Hosen *et al*. (2017) documented compositional shifts in sediment microbial communities in urbanized streams with an increase in taxa related with human activities, such as genus *Polynucleobacter* and potable water and sanitation infrastructure corrosion-linked genus *Gallionella*. Similarly, Wang *et al*. (2018) identified an increase in the expression of genes related to xenobiotic and nitrogen metabolism as well as possible fecal bacteria

1

like *Flavobacteria* and *Bacteroidia* in the Jialing River, which flows through numerous Chinese cities. These findings support the idea that urban discharges may operate as the primary selective factor to change the microbial populations. In this study, total protein (TP), NO_3^- and metals (Zn, Fe) were the most significant drivers determining the microbial community composition in the urban river. In the Changzhou region of the Yangtze River Delta Urban Agglomeration, China, analysis of fifteen rivers affected by residential sewage indicated enrichment of enteric pathogens (such as *Arcobacte*r and *Bacteroides*) and environmental pathogens (such as *Acinetobacter*, *Aeromonas*, and *Pseudomonas*) (Cui *et al*., 2019). Further, dissemination of virulence/defense (antibiotic resistance and metal resistance), degradation of organic pollutants and stress response-related genes in urbanization-impacted stream ecosystems have been reported (Chaudhary et al., 2018; Medeiros et al., 2016). According to Zhang *et al*. (2019), sewage inputs also significantly contributed to the Ganges river in India having a 13-fold higher abundance of human gut microbiome (HG) associated sequences and a 2-fold higher abundance of antibiotic resistance genes (ARGs) than other riverine ecosystems in Europe, North America, and South America. Collectively, these studies provide important insights on the impact of urbanization on bacterial community composition, giving impetus that continuous monitoring of urban aquatic microbiomes must be prioritized for the benefit of human and ecological health. Unfortunately, the impact of urbanization on stream/river microbiome remains overlooked in the sub-Saharan Africa region, despite the region undergoing rapid urbanization. Albeit few studies have shown metabolic pathways of bacterial communities found in urbanized impacted rivers/streams, sediment bacterial communities' functional profiles and potential bacterial pathogens in these lotic ecosystems also remain unclear.

On the shores of Lake Victoria, the second-largest freshwater lake in the world lies the city of Kisumu in Western Kenya. After Nairobi and Mombasa, it is the third-largest city in Kenya, with an estimated urban population of 600,000 (KNBS, 2019). The city is an example of emerging cities in Africa that are intermediary hubs and centres of agricultural and industrial production, governance, services and logistics, and estimated to habour 15% of the continent's population (Andreasen et al., 2017; Marais and Cloete, 2017). Similar to other rapidly developing cities in sub-Saharan Africa, Kisumu's insufficient infrastructure and service delivery to keep up with the city's rapid urban population expansion has led to inefficient spatial development, an increase in squatter settlements (slums), and a lack of essential facilities like drinkable water and garbage

disposal. The city is currently dealing with inadequate sewerage system, increasing solid waste generation, an overflowing open dumpsite, and pollution from uncontrolled solid waste dumping, which frequently contaminates rivers and the adjacent Lake Victoria via surface run-off (Sibanda et al., 2017). These current waste management challenges expected to rise with increase in urban population, that will further escalate the anthropogenic inputs leading to environment destruction and public health hazards. Few studies done in Kisumu City, have reported severe pollution of aquatic environment with toxic heavy metals like Cr, Cd, Ag, and Pb, including persistent organic pollutants (POPs), and pharmaceuticals like antibiotics and antiretroviral drugs (K'oreje et al., 2016; Mireji et al., 2008; Okungu et al., 2005; Onyari and Wandiga, 1989; Outa et al., 2020). In addition, fecal contamination of riverine ecosystem associated with high morbidity of diarrheal diseases within the city has been reported (Baker et al., 2018; Opisa et al., 2012). This information implies that Kisumu City's rivers and streams have long been subjected to anthropogenic pollution, becoming a sink for excess nutrients, untreated effluents, heavy metals, and xenobiotics. Little is known, however, about the diversity and structure of river microbial communities and pathogens, as well as how they respond to such intensive anthropogenic inputs. Furthermore, the extent to which bacterial communities exhibit biogeographic patterns in their distribution under urban pollution in the region remains unclear.

1.2 Statement of the Problem

Kisumu city has inadequate provision of infrastructure and services to meet its rapid human population growth with only 26 % of Kisumu's population currently connected to a sewer line, with the majority of people living in settlement areas (LVEMP, 2015). This has resulted into discharge of effluents into the rivers/streams ecosystem. However, the information on how human anthropogenic inputs impact Kisumu city's rivers and streams ecosystem is still unclear. In addition, the information on metabolic pathways of bacterial communities in Kisumu city's rivers/streams that enable them survive in such ecosystem is lacking.

Accumulating evidence show that sediment microbiome composition and biodiversity are highly sensitive to anthropogenic disturbances (Cui et al., 2019; Fisher et al., 2015; McLellan et al., 2015; Numberger et al., 2021). Particularly, the natural variability of bacterial community structure and composition is adversely affected by industrial wastewater discharges, chemical compounds from agricultural fields, and urban waste water. Unfortunately, the few articles

3

published about Kisumu City's rivers and streams have focused on chemical pollution and faecal contamination. It is therefore, not clear how changes in organic and inorganic nutrient loads affect variations in bacterial populations in Kisumu City's rivers and streams draining into Lake Victoria. In addition, water body's electrical conductivity (EC), total carbon (TC), total nitrogen (TN), soil organic matter (SOM), and metals are all impacted by rapid changes in nutrient loading (Al, Fe, Ca, Na, Mg, K and Zn). How the bacterial populations react to such environmental changes at various watershed locations is also unclear. Furthermore, no metagenomics study has been done to explain how pollution-related environmental changes in the Winam Gulf of Lake Victoria affect bacterial community structure and functioning. Therefore, a thorough analysis of the spatial bacterial distribution in the sediments of the urban streams, which flow through Kisumu City into the eastern Winam Gulf of Lake Victoria, are warranted. Furthermore, the key environmental drivers known to affect aquatic ecosystem, potential connections between composition and environmental factors need to be investigated.

1.3 Objectives of the study

1.3.1 Main objective

To assess the effect of urban pollution on the distribution of bacterial communities in Kisat and Auji streams in Kisumu.

1.3.2 Specific objectives

1. To determine the sediment bacterial community structure and diversity along the river/stream catchment impacted with varying levels of urbanization.
2. To determine the sediment microbial communities' functional profiles and the presence of potential bacterial pathogens to establish their role in ecosystem functioning and health status urban river ecosystem.
3. To identify the key environmental factors accounting for the variations in sediment bacterial community composition and structure.

1.3.3 Hypotheses

1. The community structure of sediment bacteria along the catchment of two streams (Kisat and Auji) affected by varied levels of urbanization shows no discernible difference in diversity and abundance.

2. In the catchment of two streams (Kisat and Auji) affected by various levels of urbanization, there are no appreciable differences in the predicted functional profiles of the sediment microbial communities with regard to potential pathogens.

3. There are no significant correlations between the sediment physicochemical composition and bacterial community structure sediments in the catchment of two streams (Kisat and Auji) impacted with varying levels of urbanization.

1.4 Significance of the study

Pollution of water bodies such as the catchment rivers/streams and Winam Gulf of Lake Victoria due to increased anthropogenic activities requires enhanced scientific attention, to ameliorate increased risks of depletion of its aquatic biodiversity. Specifically, urban water pollution, mostly due to the discharge of untreated domestic and industrial wastewater coupled with other anthropogenic inputs into waterways, is a growing problem in developing countries (Godoy et al., 2020; Jordaan et al., 2019; Kapembo et al., 2019; Pantha et al., 2021; Zhang et al., 2019). In this study, the observed decline of river sediment microbial quality and potential public health safety linked to the spatial distribution of heavy metal and wastewater pollution sources, illustrated this problem in the Kisumu City watershed. The two rivers under study drain into Lake Victoria, a freshwater lake that has reported widespread deterioration of water quality (Enander, 2017; Nyilitya et al., 2016; Outa et al., 2020), despite being a major source of potable water and an important fishing site in the region (Kolding et al., 2013; Nyakeya et al., 2017).

Sediment samples from polluted river sections clustered distinctively and displayed a higher bacterial diversity than non-polluted zones, strengthening the fact that sediments may serve as reservoirs of diverse bacterial populations and highlighting the need to include sediments in river monitoring (Zhang et al., 2018). Furthermore, faecal bacteria and candidate pathogens accounted for a significant portion microbiota in the pollution affected mid and lower catchment zones, implying that they have the potential to serve as important indicators for the health and status of river ecosystems. Further, the results of this study suggest that the current waste

management and sanitation provisions in Kisumu City are insufficient for the protection of public health and ecosystems. With urbanisation and rapid population growth in African cities, the challenges of urban pollution are expected to increase. Thus, immediate safe anthropogenic pollution management strategies and policies in response to the rapid urban expansion need to be undertaken to mitigate the expected severe impact on aquatic ecosystems, including disruptive effects for both humans and the environment.

CHAPTER TWO

LITERATURE REVIEW

2.1 Pollution of fresh water bodies.

All types of life require water to survive. A higher standard of living and the prevention of disease depend on the availability of potable water. Freshwater lakes are a significant supply of drinking water, so conserving and improving lake water quality is important (Mou et al., 2013). Anthropogenic activities have a substantial impact on downstream aquatic ecosystems when they occur near cities and other locations with a high population density. Waterborne pathogens are the major contributors of outbreak of infectious diseases and according to WHO (2004), waterborne pathogens are likely to cause more challenges in the future.

Over the past years, there has been improvement in municipal wastewater treatment and management of agricultural wastes. However, wastewater treatment plants (WWTPs) continue to be the main source of anthropogenic effluents entering aquatic environments (Nega et al., 2019). One of the main causes of persistent water pollution is the improper treatment of sewage or wastewater before its release into water bodies. Heavy metals and pharmaceuticals are one of the many pollutants that can be found in wastewater or sewage since some of them cannot be entirely removed by WWTPs (Nega *et al*., 2019). Human waste is another component of wastewater, which raises the level of organic nutrients in water (Nguyen, 2017). Wastewater treatment prior to their discharge to water bodies is undertaken by many countries as one of the management strategies to guarantee maintenance of water quality and make it safe for residential use. However, despite the wastewater treatment processes to reduce pathogens from sewage contaminated water, these processes are prone to system failures and of varying effectiveness leading to their presence in the final effluent and water bodies (Domingo and Edge, 2010).

2.2 Pollution of Lake Victoria.

Lake Victoria is an important freshwater resource in terms of fishing, energy, drinking and irrigation water source, and transport for the lacustrine community that is shared by Kenya (6%), Tanzania (51%) and Uganda (43%) (Table 2.1). The lake also acts as a sink for human, agricultural and industrial waste attributed to anthropogenic activities within its catchment. Human activity in catchments and along shore areas and catchment has been linked to the reduction in water quality

(Enander, 2017). In Kenya, the lake catchment area is 42460km^2 out of total catchment area of 193000km^2 in East Africa. Rivers from this region feed the lake with water, nutrients, sediments, and pollution (Okungu et al., 2005).

Table 2.1 Land catchment area in the Lake Victoria Basin (Okungu et al., 1922)

Country	Catchment area in km^2	Catchment area in %
Tanzania	84920	44
Kenya	42460	22
Uganda	30880	16
Rwanda	21230	11
Burundi	13510	7
Total	**193000**	**100**

The primary factor contributing to the deterioration of Lake Victoria's water quality is the discharge of untreated sewer and chemical wastes from urban populations, as well as microbe- and nutrient-rich runoffs primarily from agricultural lands and urban centres. Wastewater from settlement areas without sewerage systems, greywater (domestic wastewater including water from baths, showers, hand basins, and washing machines) from these households frequently drained into rivers and streams that eventually drain into Lake Victoria (LVEMP, 2015; Munala and Moirongo, 2017). The lake is supplied with other rivers such as Mara River in Tanzania and Kagera River in Uganda (Okungu *et al*., 2005), including rivers Nzoia, Gucha-Migori, Sondu Miriu, Yala and Nyando draining into Winam Gulf (Mutuku et al., 2014). In the eastern part of the Winam Gulf, Kisat and Auji river drain into the Lake Victoria through Kisumu City, Kenya. The rivers discharge effluents from domestic and industrial sources, and large volumes of eroded soil from agricultural farms rich in agrochemicals (Mutuku et al., 2014). Kisumu town among other major towns has a number of industries including brewery, tannery, textiles and fish processing industries and town sewage plant at Kisat discharge effluent into the lake without treatment (Mutuku et al., 2014). Untreated sewage ends up in the stream ecosystem through unconnected sources (Plate 1).These effluents contain organic and inorganic contents which cause proliferation of microorganisms (Mutuku et al., 2014). Kisat and Auji river travers areas characterized by different land use patterns and degree of urbanization. On that basis, the two streams can be used as

representative of typical urban river/stream ecosystem in the region and thus, can provide us with insight of what happens in other rivers/streams draining into L. Victoria.

2.3 Urbanization-linked rivers pollution

Rivers are a continuous flow of liquid water from one location such as a mountain, towards another location such as a sea, lake, or even another river (Nguyen, 2017). Anthropogenic activities (e.g., mining, urban and rural settlements, agriculture and sewage works) have been increasing overtime resulting into disturbances on river systems (Chen et al., 2022; Jordaan et al., 2019). Due to excessive pollution, including microorganisms, the water quality of the river has degraded markedly and continues to deteriorate all over the world. Riverine ecosystems are known to be easily affected by external pollution and because of the flow nature of river water, thus, pollution spread faster into the entire river ecosystems (Han et al., 2021; Jordaan et al., 2019). These contaminations may cause undesirable changes in the physical, chemical, and biological characteristics of water sources, which may have an impact on the aquatic ecosystem and on people's health. Therefore, various organizations have established acceptable levels for discharge of industrial, domestic and municipal wastewater into public water. The discharge includes heavy metals in soils, wastewater, and drinking water. In this study, the recommended limits by World Health Organization (WHO), United States Department of Agriculture Natural Resource Conservation Service (USDA NRCS), National Environment Management Authority (NEMA), Kenya Bureau of Standards (KEBS), and Water Service Regulatory Board (WASREB) (Table 2.2) were used to evaluate the quality of river ecosystem.

In the recent years, pollution of urban rivers has been on rise due to increase in human population, industrialization and rapid expansion of urban infrastructure which mainly has resulted into deposition of solid and liquid waste materials into river waters (Saraceno et al., 2021; Wang et al., 2020a; Zinabu et al., 2019). Pollution of urban rivers have increased mainly due to rapid urban population growth as compared to industrial and agricultural activities (Li et al., 2022; Wang et al., 2020b). Pollutants in urban rivers are mainly from domestic and industrial effluents, agricultural pollution sources and other urban business activities such as car washing among others. Domestic wastewater and industries account for the majority of organic and inorganic contaminants. Grey water (domestic wastewater that includes water from baths, showers, hand basins, and washing machines) from households are often drained into river and streams and are

known to contain organic contaminants mainly from processing of vegetables, oils and fats, dairy products, meat, tanning, and paper, as well as synthetic detergents (Bodnar *et al*., 2014; Kayaga *et al*., 2005; O'Toole *et al*., 2012).

Table 2.2 Guideline values for discharge into public water (The Environmental Management and Co-ordination (Water Quality) Regulations, 2006)

Parameter	unit	Guideline value
Cadmium (Cd), max	mg/l	0.01
Copper (Cu), max	mg/l	1.0
Mercury (Hg), max	mg/l	0.005
Total Chromium (Cr), max	mg/l	2.0
Lead (Pb), max	mg/l	1.0
Zinc (Zn), max	mg/l	0.5
Dissolved Iron	mg/l	10
Dissolved Manganese	mg/l	10
Ammonia Nitrogen	mg/l	20
Total Phosphorous	mg/l	2
Total Nitrogen	mg/l	2

Organic pollutants contain hydrocarbons e.g., oil, polyaromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs). Polyaromatic hydrocarbon substances are toxic, persistent, carcinogenic and a mutagen in the environment (Bacosa and Inoue, 2020). The dissolved organic material is broken down by microorganisms, and it is then made available to the upper trophic levels of the food chain. Organic materials made of hydrocarbons e.g. oil, polyaromatic hydrocarbons (PAHs), when decomposed by microorganisms leads to reduced level of oxygen in river environment (Nguyen, 2017). Given these dramatic effects, there is an urgent need to monitor diffuse and non-diffuse sources of river water pollution and to assess the status of these pollution sources for both immediate and future mitigations.

Inorganic substances are also major pollutants in river ecosystem. They include heavy metals such as Cadmium(Cd), arsenic (As), lead (Pb), silver (Ag), mercury(Hg), Zinc(Zn), copper

(Cu), chromium (Cr), Iron (Fe), boron (Bo) and platinum (Pt), some of which are non-biodegradable (Ancion et al., 2010; Lee et al., 2005; Nguyen, 2017). According to Ali et al., (2021) heavy metals are defined as elements with an atomic number more than 20 and metallic properties like ductility. Heavy metals occur in the soil in soluble form and in combined state (Chiroma et al., 2014). In aquatic environments, heavy metals are naturally extremely soluble and poisonous in mixed or elemental forms (Kinuthia et al., 2020).

Increased production of wastes with high concentrations of heavy metals is the result of anthropogenic activity. Some heavy metals are essential for normal biological functions but their elevated concentrations above what is considered safe for biological processes and the environment could be harmful (Fashola et al., 2016; Kinuthia et al., 2020). The negative health effects include potential carcinogens, damage to the brain, lungs, and kidneys, harm to growing fetuses, high blood pressure or heart rate, eye discomfort, and skin rashes (Kinuthia et al., 2020). Heavy metals that are toxic at elevated levels include Cadmium (Cd), lead (Pb), mercury (Hg), Zinc (Zn), silver (Ag), copper (Cu), chromium (Cr), arsenic (As), Iron (Fe), Nickel (Ni), Thallium (Tl), boron (Bo) and platinum (Pt) (Fashola et al., 2016; Kinuthia et al., 2020; Nguyen, 2017). Pollutants from agro based and other chemical industries have high concentration of heavy metals (Mutuku et al., 2014). Pollutants from industries include heavy metals like mercury, cadmium, chromium, lead, copper, arsenic, and zinc that are discharged into the waterways (Iloms et al., 2020). When these pollutants accumulate into toxic levels, they cause deterioration of river water quality and are very toxic to aquatic organisms (Wang et al., 2020a).

2.4 Urban pollution in Kisumu City

Kisumu City is an example of rapidly growing intermediary urban centres in sub-saharan. The city has inadequate provision of infrastructure and services to meet its rapid human population growth. For example, only 26% of the Kisumu population is connected to sewer line with the bulk of population in the settlement areas including Obunga and Nyalenda slums, Mamboleo, Nyawita, Kibuye and Manyatta not connected. It is estimated that only about 11,000m^3/day of the sewage generated reaches municipal wastewater treatment facilities against estimate of sewage generated of 34,000m^3 (LVEMP, 2015). However, despite low feeds, the two wastewater treatment plants are also prone to system failure due to poor maintenance, overloading and inefficiency of certain components. This suggests that inadequately processed sewage from the treatment plants or raw

sewage from unrelated sources is frequently released into the streams in Kisumu and Lake Victoria (Plate 2.1). Despite widespread use of septic systems for toilet wastewater management in the settlement areas not connected to sewerage system, grey water (domestic wastewater that includes water from baths, showers, hand basins, and washing machines) from these households are often drained into river and streams that eventually drains into Lake Victoria (Kayaga et al., 2005).

Plate 2.1. Untreated sewage discharged into River Kisat within Obunga

In addition, the city has a number of industries including a brewery, tannery, textiles and fish processing industries which discharge treated and untreated effluents into the streams. Other forms of pollution, such as solid wastes, are challenging to regulate. In addition, it is common for roadways to be improperly cleaned and for oil spills from carwashes and auto repair shops to contaminate the soil and rivers. In addition to dealing with liquid wastes, Kisumu is currently dealing with rising solid waste creation, an overflowing open dumpsite, and contamination from the careless disposal of solid wastes, which frequently contaminates nearby rivers and Lake Victoria via surface run-off (Sibanda et al., 2017). According to census projections, Kisumu's population would grow by 2.8% annually, placing further strain on the city's infrastructure and provision of essential services (Munala and Moirongo, 2017). With rapid population growth

coupled with changing lifestyles and consumption patterns, waste generation is likely to increase further amplifying the current waste management challenges. This will further escalate the anthropogenic inputs leading to environment destruction and public health hazards. Therefore, there is an urgent need for effective strategies to address both solid and liquid waste management in Kisumu City. Furthermore, studies on assessment of urban river water quality and microbiota in Kisumu City can play a vital role in the management of the environment as the city has a distinct urban form and land use pattern that affects the kind of wastes created and pollution tendency on the rivers within its catchment.

The few studies that have been published on the rivers and streams that traverse Kisumu City have, to date, focused on chemical pollution (Mireji et al., 2008; Okungu et al., 2005; Onyari and Wandiga, 1989; Outa et al., 2020) and fecal contamination (Baker et al., 2018; Opisa et al., 2012). These studies have reported severe pollution of toxic heavy metals like Cr, Cd, Ag, and Pb, including persistent organic pollutants (POPs), and pharmaceuticals like antibiotics and antiretroviral drugs (K'oreje et al., 2016; Kimosop et al., 2016), which pose serious risks to both the environment and human health. Thus, it can be postulated that streams and rivers that flow through Kisumu City have been a sink for the excessive nutrients, untreated effluents, faecal microbiota, heavy metals and xenobiotics due to long exposure to anthropogenic pollution.

2.5 Bacterial community structure in riverine ecosystems

2.5.1 Bacteria in natural environments

The makeup of the bacterial community in natural habitats has been the subject of numerous previous studies. The bacterial community composition is influenced by the environment, which has an impact on the microbial population. When studying bacterial community composition and diversity in any natural ecosystem, the following questions are commonly addressed. What bacteria are present and in what proportions? (Nguyen, 2017), how do they live in the given natural environment? (Haruta & Kanno, 2015). To survive in any natural environment, bacteria must adapt to overcome various environmental stresses (Haruta & Kanno, 2015). The adaptive strategies may include changes in the cellular membranes, cysts and spores formation, synthesis of molecules for relieving stresses, chemotactic behavior and programmed cell death (Haruta and Kanno, 2015). For example, it has been reported that *Methylobacterium* may survive and proliferate in environments with little nutrition supply and sporadic exposure to

cleaning supplies (Zhao, Huang, Zhu & Chai, 2013). By altering their growth rate, cell localisation, and increasing cell surface lipophilicity, *Rhodococcus sp.* can become resistant to an organic solvent (Haruta & Kanno, 2015). In soil sediments, it has been revealed that the phyla *Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, Acidobacteria,* and *Verrucomicrobia* predominate (Nguyen, 2017). At genera level, *Norcardia, Agrobacterium, Arthrobacter, Streptomyces, Alcaligenes, Pseudomonas* and *Bacillus* are found in soil environment in relatively high proportions (Nguyen, 2017). However, In Kisumu city rivers/streams metagenomics profiling based on biological functions of bacterial communities' distribution remains unsolved.

2.5.2 Bacteria in polluted environments

Depending on types of pollutants and environments, studies reveal that different bacterial taxonomic groups are found in different polluted habitats and can be used as indicator of pollution. Discharge of sewage into freshwater sources may endanger public health since sewage is thought to be a habitat for a variety of human bacterial pathogens. Waste water contain numerous pollutants originating from various effluents such as pharmaceutical effluents, textile effluents and dye effluents among others (Aneyo et al., 2016). Several bacteria such as *Bacillus cereus, Bacillus subtilis Salmonella typhi* and *Klebsiella pneumonia* have been isolated from both textile and pharmaceutical effluents (Aneyo et al., 2016). Bacteria such as *Arcobacter butzleria*, *Aeromonas hydrophila* and *Klebsiella pneumonia* are predominantly common in raw sewage (Lu *et al*., 2015). Members of genus *Burkholderia*, are abundant in oil polluted habitats since they are involved in oil degradation in soil. *Burkholderia cepacia* and *Pseudomonas sp* has been found to be the main bacteria in the degradation of heavy oil which is a major products from crude oil, utilizing it as the sole source of carbon (Bacosa and Inoue, 2020).

In soil, *Actinobacteria* have been found to decompose diesel and fuel oil (Nguyen, 2017). Heavy oil contains significant amounts of polycyclic aromatic hydrocarbons (PAHs), which are harmful to the environment (Bacosa and Inoue, 2020). PAHs are broken down in soil by bacteria such as the Actinobacterial genus *Mycobacterium,* the Proteobacterial genus *Sphingomonas,* the γ-Proteobacterial *genus Pseudomonas,* and the β-Proteobacterial genus *Burkholderia.* In urban river sediments that are plentiful in organic compounds, petroleum byproducts, and PAHs, Cyanobacteria and Bacteroidetes are particularly prevalent (Nguyen, 2017). Various bacteria can be used to predict type of heavy metal contamination since they respond differently to heavy metal concentrations. Due to their capacity to adapt to a variety of settings and their capacity to

biodegrade heavy metals, *Proteobacteria* have also been discovered to be present in large quantities in soils that are contaminated with heavy metals (Luo and Ma, 2018). Studies have also revealed that members of the *α-Proteobacteria* can be correlated to elevated levels of HMs. For instance, *α-Proteobacteria* positively correlated with Zn and Cd, whereas *Chloroflexi* were favorably correlated with high levels of Hg (Nguyen, 2017). *Nitrospira, Polycyclovorans* can be used to predict pollution with Cu, *Ramlibacter* to predict Zn contamination while *Steroidobacter* can be used as indicator of pollution with Cd (Luo and Ma, 2018).

2.5.3 Impact of pollution on river sediment microbiome

River ecosystems are vital resource for a variety of human uses and are incorporated into biogeochemical cycles (Köchling et al., 2017). Rivers and streams that flow through cities are constantly under threat from a variety of environmental stressors because they serve as important sinks for waste, sewage, and storm water run-off. This pollution has the potential to change the structure and composition of rivers and streams, affecting their geomorphology, physicochemical properties, eukaryotic and prokaryotic community structure, function, and health (Beattie et al., 2020; Chaudhary et al., 2018; Fisher et al., 2015). Contaminations of river sediments are known to exist for a longer period of time than in river water. Therefore, sediments serve as primary sites for the accumulation and attachment of microorganisms and chemical pollutants in riverine ecosystems. Consequently, these habitats are more stable and bacterial residence time is longer, thus sediment microbial communities are more strongly linked to environmental conditions within the river/stream (Beattie et al., 2020).

A major change in bacterial community composition has been observed in freshwater systems due to pollution (Jiang *et al.*, 2006). Bacterial diversity in freshwater ecosystem sediments are enriched with members of *Acidobacteria*, *Verrucomicrobia*, *Alphaproteobacteria* and *Betaproteobacteria* (Wang et al., 2012). In addition, saprophytic microbes such as *Bacteroidetes* and *Firmicutes* are also common in freshwater systems (Wang et al., 2012). However, polluted rivers are enriched in bacterial families such as *Aeromonadaceae, Bacillaceae, Bacteroidaceae, Bacteroidales, Clostridiaceae, Enterobacteriaceae, Maraxellaceae, Peptostreptococcaceae* and *Porphyromonadaceae. Enterobacteriaceae, Bacillaceae* and *Clostridiaceae* with some taxa being human pathogens transmittable through water (Mukherjee et al., 2016). Other phyla found in polluted freshwater include *Deltaproteobacteria, Epsilonbacteria, Acidobacteria, Fusobacteria, Flavobacteria* and *Fibrobacteria* (Jordaan and Bezuidenhout, 2013).

Overall, urban-linked pollution impacts the riverine microbiome due to: i) increased loads of organic wastes (attributed to domestic and industrial wastewater, including greywater pollution); ii) inorganic wastes (heavy metals); and iii) contamination with fecal and human associated microbiota from untreated dometic wastewater and grey water contamination. River system especially sediment are complex primary sites densely settled by various microbial communities (Beattie *et al.*, 2020). Sediment microbiome composition and biodiversity are highly sensitive to aforementioned anthropogenic disturbances (McLellan et al., 2015), and can provide significant insight into the degree of human impacts on stream ecosystems, the presence of human health risks, or valuable clues on the short- and long-term ecosystem alteration (Cui et al., 2019; Fisher et al., 2015; McLellan et al., 2015; Numberger et al., 2021). Under conditions of urban pollution, bacterial communities in rivers and streams play key roles in biogeochemical cycles and are responsible for breaking down organic material and recycling nutrients (Xie *et al.*, 2022). Previous research has also demonstrated the interactions between pollutants such as heavy metals and the river sediment microbiome. For instance, it has been reported that *Bacteroidetes, Acidobacteria, and Proteobacteria* are resistant of Zn, Pb, and Cd contamination (Medeiros et al., 2016; Saraceno et al., 2021; Wang et al., 2018; Zhao et al., 2020). Hosen et al (Hosen et al., 2017) reported compositional shifts in sediment bacterial communities in urbanized streams with an increase in taxa associated with human activity such as genus *Polynucleobacter* and corrosion of water distribution systems-linked genus *Gallionella* in 8 rivers within Parkers Creek watershed in Michigan, USA. Similarly, Wang et al (Wang et al., 2018) reported an increasing trend of potential faecal bacteria such as *Flavobacteria* and *Bacteroidia*, as well as genes associated with xenobiotics and nitrogen metabolism, in the Jialing River, which flows through many Chinese cities.

Urbanization is also linked with the release of large quantities of greywater into environment. Greywater is the term for residential wastewater, which makes up between 50 and 80 percent of all domestic wastewater and often comprises water from bathrooms, hand basins, showers, and washing machines (Gross, 2007; Jefferson et al., 2000). Pathogenic bacteria in greywater pose potential health risk to humans and is currently a critical issue (Gross, 2007). Even though grey water does not include toilet waste, presence of fecal coliform have been found to be high due to washing diapers or children in shower or bath or hand basin (James et al., 2016). Fecal enterococci, fecal streptococci, *Klebsiella pneumoniae*, *Salmonella* spp., *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Clostridia* have all been identified as

bacterial pathogens in greywater (Jefferson, Palmer, Jeffrey, Stuetz, & Judd, 1999; Katukiza, Ronteltap, Niwagaba, Kansiime, & Lens, 2015; Toole *et al*., 2013). However, among the above bacterial pathogens, *E. coli* has been the most frequently detected and widely distributed pathogen (Birks and Hills, 2007; Bodnar et al., 2014). Bathrooms, which have rapid water flow, drying conditions, low nutrition levels, and sporadic exposure to cleaning products, are another location where *Methylobacterium* predominate (Zhao *et al*., 2013). Despite their important role in river ecosystems, the assessment of riverine biological diversity and composition of microbial communities, still remains partly addressed due to methodological constraints (Köchling *et al*., 2017). Similarly, in Sub-Saharan Africa, despite rapid urbanisation, the impact of urbanisation on stream/river microbiome remains unexplored. In the study area, little is known about the distribution and structure of the microbial population and pathogens in stream sediments, as well as how they respond to intense anthropogenic inputs. Because each city has a distinct urban structure and land use pattern that influences the type of waste generated and pollution tendency on the rivers that flow through it, assessing Kisumu urban rivers and streams water quality and microbiome can play an important role in environmental management and thus warrants a study.

2.6 Health risks associated with bacterial contamination of water

The spread of water-borne diseases driven by bacteria in polluted waters is a serious global concern. Millions of people in Africa, for instance, suffer water-borne diseases as a result of using contaminated water. More than 27% of cases of gastrointestinal disease (GI) are caused by *Shigella* spp., whereas only 10.99% are caused by *Cryptosporidium parvum*. *Naegleria fowleri*, *Escherichia. coli* 0157:H7, and *Schistosoma* spp. are responsible for about 16.84, 12.63 , and 7.37 %, respectively, of outbreaks of gastrointestinal infections. (Birks & Hills, 2007; Bodnar *et al*., 2014; Jefferson *et al*., 1999; Katukiza *et al*., 2015; Toole *et al*., 2013.; Winward *et al*., 2007) and can cause serious human diseases (Toole *et al*., 2013). *Vibrio cholera* and *Salmonella* are also responsible for many water-borne disease outbreaks. Around 3.4 million people died from water-related diseases in 2014, primarily children (WHO, 2014). Each day, polluted water claims the lives of about 4,000 youngsters (UNICEF, 2014). WHO estimates that more than 2.6 billion people lack access to safe water, which results in 2.2 million annual deaths, 1.4 million of which are children (WHO, 2011). In Kisumu City, there are several reports that urban neighborhoods such as informal settlements and streams are heavily contaminated by many enteric viral, bacterial,

protozoan, and helminthic enteric pathogen species (Baker et al., 2018), with records from health facilities showing that diarrheal diseases are among the top causes of morbidity. High levels of environmental contamination, often associated with improper waste and excreta management, is widespread among informal settlements within Kisumu City that contaminates public water sources (Opisa et al., 2012; Rochelle-Newall et al., 2015).

2.7 Tools used in studying environmental bacteria

2.7.1 The 16S rDNA gene-based bacterial identification

Previous studies have attempted to assess riverine biological diversity and microbial community composition. However, this still remains partly addressed due to selective bias introduced by the available methodology, were mainly based on culture based techniques and limited molecular methods (Köchling *et al*., 2017). However, in the recent years, this is being resolved by use of Illumina and 454-pyrosequencing of the 16S gene and these has provided far more detailing information on riverine microbial communities (Köchling *et al*., 2017). These advance techniques utilize the 16S rDNA gene, a roughly 1550 bp long, highly conserved DNA sequence that is a component of the ribosomal small subunit. The 16S rRNA gene possesses nine hypervariable regions (V1-V9) that show significant sequence diversity among various bacteria and can be exploited for species identity (Chakravorty et al., 2007). Greater higher taxa can be identified by more conservative regions, whilst genus or species can be identified by more hypervariable ones. For example, hypervariable regions V1-V3 or V3-V5 are usually used to study 16S rDNA for identification of bacteria at most taxonomic levels for example in human microbiome, while regions V2-V3 is suitable to differentiate bacteria at genus level (Johnson *et al*., 2019).

2.7.2 High-throughput metagenomic sequencing (HTS)

DNA sequencing has become one of the most important platforms for the study of biological systems today (Reuter et al., 2015). When selecting DNA sequencing platform, there are at least four major options for metagenomic studies including dideoxy sequencing (Sanger), pyrosequencing (454 – Roche), SOLiD™ (Applied Biosystems), and Illumina® (formerly known as Solexa).

2.7.2.1 Dideoxy sequencing (Sanger)

Sequence determination is most commonly performed using di-deoxy chain termination technology (Reuter et al., 2015). Sanger sequencing is also known as chain termination sequencing, has become the most widely used method of DNA sequencing since its advent in 1977 by Frederick Sanger and colleagues and still is in use after more than 29 years (Olson et al., 2009). Despite all the advantages, there are limitations in this method, which could be complemented with other techniques. In Sanger sequencing, DNA fragments of varying lengths are synthesized by incorporating both nucleotides and dideoxy terminators (Olson et al., 2009).

2.7.2.2 Pyrosequencing (454 – Roche)

Pyrosequencing is the first alternative to the conventional Sanger method for de novo DNA sequencing (Reuter et al., 2015). Roche 454 platform is a high-throughput sequencer used as a method to detect the hidden diversity of microorganisms. It is based on sequencing by synthesis and relies on the real-time detection of inorganic pyrophosphate (PPi) released on successful incorporation of nucleotides during DNA synthesis (Olson et al., 2009). It employs a series of four enzymes to accurately detect nucleic acid sequences during the synthesis (Reuter et al., 2015). Roche 454 provided relatively long sequences, suitable for amplicon sequencing of specific target regions (Hirai et al., 2017). Pyrosequencing has the potential advantages of accuracy, flexibility, parallel processing and can be easily automated (Lightbody et al., 2019; Reuter et al., 2015).

2.7.2.3 Illumina® (formerly known as Solexa)

Illumina MiSeq is integrated next generation sequencing instrument that uses reversible terminator sequencing by synthesis technology to provide end-to-end sequencing solution (Bomar, Maltz, Colston, & Graf, 2011). Illumina MiSeq is becoming more common and most widely used next generation sequencing platforms in metagenomics and gene expression studies since it can more than 10 times the number of sequence reads of Roche 454 at a lower cost (Caporaso et al., 2012; Kozich et al., 2013). The MiSeq instrument is one of the smallest benchtop sequencers that can perform onboard cluster generation, amplification, genomic DNA sequencing and data analysis, including base calling, alignment and variant calling, in a single run (Kozich et al., 2013).

The newer methods, including Roche/454 and Illumina/Solexa, are referred to as next-generation sequencing. Next-generation sequencing technologies have revolutionized the analysis of microbial communities in diverse environments (Fouad, 2012). One of the advantages of next-

generation sequencing is the ability to produce an enormous volume of data cheaply. Next generation sequencing (NGS) technology have led to source identification based on the comparative characterization of the entire microbial communities of environmental samples and pollution sources (Ahmed et al., 2015).

2.7.3 Microbial source tracking

Microbial source tracking identify the potential origin of microbiome compositional structure (Shenhav et al., 2021). Among the different MST-based approaches, the marker gene approach, which relies on measurement of host source-associated DNA sequences by polymerase chain reaction (PCR)-based technologies, is among the most utilized (Harwood et al., 2013). These methods differ from the single marker MST-based approaches by their ability to characterize thousands of sequences in each sample (Sogin et al., 2006). However, the current methods for microbial source tracking have some limitations. Earlier methods were limited to contamination, focusing on detection only specific, predetermined contaminating species. The current methods of microbial source tracking often lack a proper probabilistic framework or depend on the identification of indicator species, whose abundance reflects a specific environmental condition (Shenhav et al., 2021).

Several studies have explored the use of metabarcoding to expand our knowledge about the river microbial communities as well as their physiological functions in response to changes in environmental parameters. For instance, a study conducted in intermittent peri-urban Chaudanne River showed that Illumina and 454-pyrosequencing of the 16S gene appeared more sensitive at tracking bacterial community changes than microbial source tracking (MST) markers. A study in Wonderfonteinspruit catchment area, South Africa used 16S rRNA gene high-throughput sequencing to characterize bacterial communities and their response to various contaminant sources (Jordaan et al., 2019). These advances in sequencing technology are revolutionizing our understanding of the roles microbes play in health and disease, biogeochemical cycling and their response to various contaminations (Knights et al., 2011). Although previous studies have used various sequencing technologies, high throughput sequencing rarely has been used in studies of bacterial communities in urban-impacted rivers/streams, particularly in sub-Saharan Africa.

2.7.4 Use of metagenomics in studying environmental microbiome

Metagenomics involves direct analysis of collective genomes of biological communities found within environmental samples (Xu, 2006). Metagenomics approach was originally used to detect unculturable bacteria found in sediments (Jordaan and Bezuidenhout, 2013). Metagenomics provides extensive information on microbial community structure and composition (Jordaan and Bezuidenhout, 2013). Metagenomics involve cloning of DNA fragments, followed by sequencing and finally functional analysis of the cloned segments (Xu, 2006). Only a small portion of bacteria could be cultivated on selective media, which has long hampered research on wild bacterial populations (Bryanskaya *et al.*, 2016). Metagenomics answers two questions, "who is there? " and " what might they be doing? ". Bacterial identification through sequencing of maker genes and whole genome sequencing answer the question " what is there? " (Nguyen, 2017). Using a metagenomics method, natural bacterial populations of benthic and pelagic bacteria can be examined by directly sequencing their 16S rDNA genes. A much more comprehensive description, such as those of bacterial biocatalysts or enzymes, is provided by metagenomics, which enables access to the functional gene composition of bacterial populations (Mukherjee et al., 2017). Metagenomics studies have been useful in elucidating microbial communities found in toxic environments. A metagenomics study conducted in Jialing River, which flows through numerous Chinese cities showed high abundance of *Flavobacteria* and *Bacteroidia* which have been associated with fecal pollution (Wang *et al.*, 2018). *Proteobacteria, Bacteroidetes, Actinobacteria, Chloroflexi* and *Acidobacteria* were found to be dominant in a tropical river highly impacted by urbanization activities in northeastern Brazil (Köchling *et al.*, 2017) Therefore, metagenomics can be used to provide details of interactions between bacterial communities and their natural and anthropogenic alter environments. Despite the numerous studies conducted to assess the impact of urbanization in many countries, limited efforts have been put on investigating how microbial composition and structure are affected by varied levels of urbanization in developing countries.

In this study, Illumina MiSeq sequencing platform for 16S rRNA amplicon analysis was used to allows assessment of mixed stream sediment microbial communities by directly sequencing DNA from environmental samples. This approach provided a picture of the diversity and microbiome structure present in the environment, and it enables studies to understand how microbial diversity is modulated in response to environmental or anthropogenic impacts. This method has been found to be particularly appropriate for assessing the taxonomic and functional

microbial diversity in river sediment biome and monitoring community changes over space and time (Beattie et al., 2020; Brito et al., 2013; Chaudhary et al., 2018; Wang et al., 2018; Yao et al., 2017).

2.7.5 Predictive metagenomic function profiling

Technologies used for revealing the functional potential of microbial communities are hindered by high costs and skills barrier necessary to generate and interpret data (Wilkinson et al., 2018). One of the tools that have been developed to address this is Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt). This bioinformatics tool was developed to predict functional potential of microbial communities from 16S rRNA studies with little extra work or cost, but its accuracy relies on the availability of complete sequenced genomes of representative organisms from the community being investigated (Sun et al., 2019; Wilkinson et al., 2018). Several studies have been conducted to address the metabolic functions of microbial communities in river ecosystems impacted by varied levels of anthropogenic inputs. A study conducted along the Apatlaco River, Mexico, reported differential distribution of biodegradative functions related to the degradation of xenobiotics. The study identified metabolic potential related to bioremediation activities, of genera *Thiomonas, Polaromonas, Pedobacter, Myroides, Pseudomonas, Acinetobacter, Aeromonas* and *Thauera* (Breton-deval et al., 2020). Another study conducted in Wonderfonteinspruit Catchment Area (WCA), South Africa, reported higher relative abundances of genes involved in xenobiotic degradation (Jordaan et al., 2019). Zhang et al (2019) also found that sewage inputs increased the abundance of human gut (HG) microbiome associated sequences by 13-fold and antibiotic resistance genes (ARGs) by 2-fold in the Ganges River in India. Despite several studies on functional profiles of bacterial communities in rivers impacted by human activities, information of functional profiles of bacterial communities in rivers in developing countries still remains unclear. The information on metabolic pathways of bacterial communities in urban impacted rivers to establish their role in the ecosystem can play an important role in environmental management.

2.8 Organization of the thesis

This research was conducted to determine the diversity and abundance of sediment bacterial communities along the two urban river/stream impacted by varying levels of

urbanisation, using 16S rRNA gene-based Illumina MiSeq sequencing (**Objective 1**). In addition, to understand the functional and metabolic potential of the microbial communities in Kisat and Auji streams to establish their role in the ecosystem, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) algorithm was used to predict functional profiles of bacterial communities by assigning metagenomics reads to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (**Objective 2**). Finally, multivariate analyses techniques such as redundancy analysis (RDA) and linear regression was used toidentify the key environmental factors that influenced compositional variations in stream sediment bacterial communities (**Objective 3**).

The subsequent chapters of this thesis consist of a consolidated theme (based on the three objectives), which are linked through the 16S rRNA gene-based reconstruction the taxonomic and functional diversity of the bacterial communities as impacted with varying levels of pollution (nutrient factors and heavy metals content). **Chapter 3** outlines the material and methods used in the study including the statistical data analysis and visualization. **Chapter 4** and **5,** further provides the results and describes the diversity of these stream sediment bacterial community, and how their taxonomy and functional attributes are affected by urban pollution gradient. The final chapter (**Chapter 6**) synthesizes the research contributions of my thesis and gives suggestions for future research directions.

The output of this study has been submitted for publication in internationally refereed journal:

1. **Odhiambo, K.,** Henry JO Ogola, Memory Tekere and Grace N. Ijoma. ***Contribution of pollution gradient to the sediment microbiome and potential pathogens in urban streams draining into Lake Victoria (Kenya).*** Environmental Science and Pollution Research (ESPR) (*Accepted*).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Description of study area

This study was conducted between September and December 2019, on two rivers (Kisat and Auji) that drain into the eastern part of the Winam Gulf of Lake Victoria through Kisumu City, Kenya (Figure 3.1). The catchment area of the streams is within latitude $00^o6'$ N and longitude 34° 45' E, covering 417 km^2 (Figure 1). Kisat and Auji rivers cover a total length of ~7.0 and 10km, respectively, through Kisumu City, traversing areas characterized by different land use patterns and degree of urbanization.
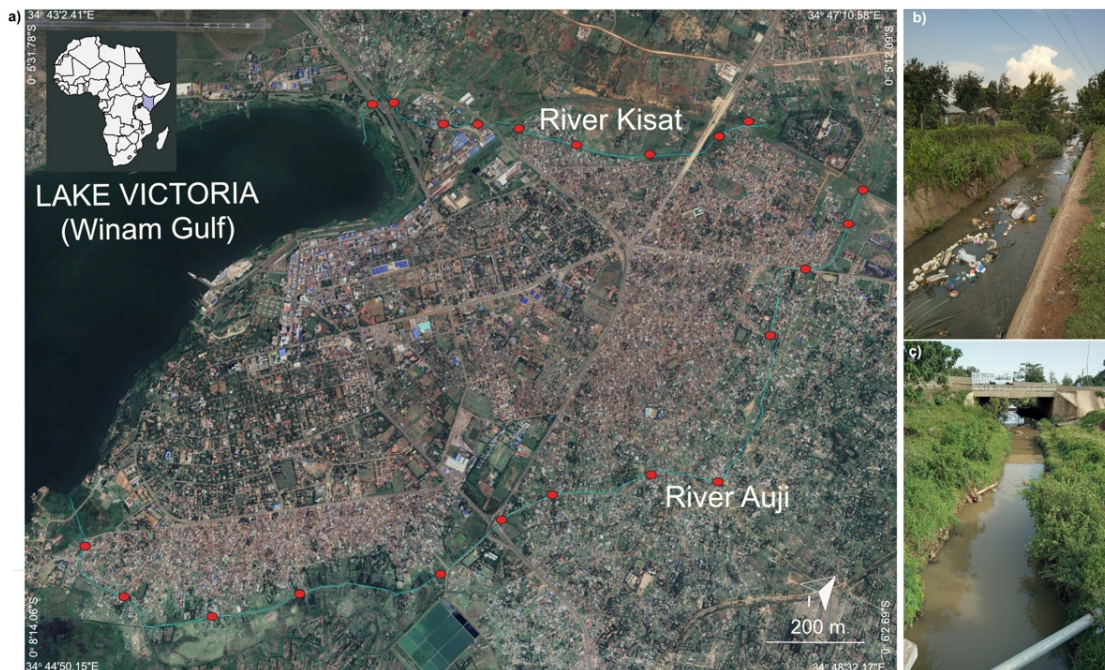
Figure 3.1 A map of the study site. a) A Google Earth map of Kisumu City showing two rivers under study (Kisat and Auji) draining into Winam Gulf of Lake Victoria, Kenya. The sampling points are denoted as the red dots. Inset is map of Africa and location of Kenya. b, and c) shows the pictures of the polluted sections of River Kisat and Auji, respectively

On the basis of land use patterns and degree of urbanization, the catchment zones of both streams were defined as either highly impacted (Mid and Lower Catchment), and one relatively less impacted (Upper Catchment). The Upper Catchment included the area covering the streams origin, either as subterranean or surface groundwater (spring), to the peri urban of the city, largely characterized by a land use pattern comprising of >70% agricultural area (mainly small-scale maize farms and large-scale sugar plantations) and < 30% developed urban area. The Mid Catchment zone is characterized by >70% developed urban area consisting of planned residential areas and informal settlements (slums) with sparse light industries. In this zone, less 30% of the population is served with the sewerage system. Finally, the lower catchment zone boarders the lake front and includes the heavily built central business district characterized by heavy concentration of industries (such as a brewery, tannery, textiles, steelworks and fish processing industries), upmarket residential settlements and motor vehicle garages. The zone has two wastewater treatment plants one draining in Kisat and Auji River, respectively.

3.2 Experimental design and sampling

The study adopted stratified purposive sampling design, where the catchment of the two streams (Auji and Kisat) were initially stratified into three strata, as either highly impacted (Mid and Lower Catchment), and one relatively less impacted (Upper Catchment), on the basis of land use patterns and degree of urbanization as described in subsection 3.1. Based on the nature of the stream and terrain, simple random sampling method was used select sampling points within each catchment zone. Specifically, a transect was laid along the river and both sediment and water samples were collected from at least three sites (0.5 Km apart) within each catchment zone of each stream. Based on this criterion nine and thirteen samples were collected from Kisat and Auji stream respectively. Based on catchment zones, six, ten and six samples were collected for the Upper, Mid and Lower catchments, respectively as shown in Figure 3.1.

Sediments were collected at the bottom of the river (0-10 cm depth) using dredge sampler (Kajak, KC-Denmark) based on a multi-point mixed sampling method (Jacquiod et al., 2018). For every sampling site, 6 replicate samples were collected with an area of 2 m x 2 m and then mixed into one composite sample (the sediment, approximately 250 g) which was placed in pre-sterilized aluminum boxes and stored in ice bags at low temperatures. Water samples were collected in triplicate approximately 20 cm below the water surface at each sampling site, in 500 mL sterile

polyethylene bottles, stored in ice bags and transported quickly to Jaramogi Oginga Odinga University of Science and Technology laboratory within 12 hours for analysis. In the laboratory, sediment samples were divided into two groups for molecular and physicochemical parameters analyses. For metagenomic experiments, samples were stored at -20°C until analysis.

3.3 Laboratory analytical methods, data visualization and statistical analyses

3.3.1 Determining the sediment bacterial community structure and diversity using the targeted 16S rRNA gene amplicon Illumina MiSeq sequencing

3.3.1.1 Genomic DNA extraction

Using the Fecal/SoilTotal DNA^{TM} extraction kit (Zymo Research Corporation, CA, USA), genomic DNA was extracted in triplicate from the 0.25 g sediment samples and stored at -80°C before further analysis. Briefly, 0.25 g of the sediment soil was initially mixed with 5 mL phosphate-buffered saline (PBS, pH 7.4). The mixture was agitated by vortexing and allowed to stand for 15 minutes at room temperature to dislodge the bacterial cells adhering to the solid soil particles. An aliquot of 400 μL resultant supernatant was used for DNA extraction using the extraction kit according to manufacturer's instructions. To account for biases introduced during the DNA extraction step, genomic DNA was extracted in triplicates for each sediment sample and pooled together before use for downstream analysis. The quantity and the quality of extracted DNA were checked on Biodrop μLite spectrophotometer (Biochrom Ltd, Cambridge, UK) and agarose gel electrophoresis, respectively. In this study, only high-quality genomic DNA (20–150 ng/L and $A_{260}:A_{280} = 1.8$–2.0) was used for downstream PCR and sequencing analyses.

3.3.1.2 Library preparation

Following the procedure outlined by Ogola et al. (2021), initial PCR amplification of the whole variable region of bacterial 16S rRNA of genomic DNA was done using the universal primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492 (5'-TACGGYTACCTTGTTACGACTT-3'). Each 25 μL reaction volume contained 0.5 μM of each primer, 1X OneTaq® Hot Start Master Mix (New England Biolabs, Ipswich, MA, USA) and 20 ng DNA. PCR was performed under following cycling conditions (95 °C, 5 min; 32 x [95 °C, 1 min; 55 °C, 1 min; 72 °C, 1 min]; 72 °C, 7 min; 4 °C, ∞), and resultant amplicons checked on a 1.5% agarose gel. A second PCR amplification to cover the bacterial 16S rRNA V1-V3

hypervariable region was performed using 27 F (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG- 3′) and 518R (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGT ATCTAATCC-3′) primer pairs, fused with MiSeq adapters and heterogeneity spacers compatible with Illumina indexes for multiplex sequencing. The resulting amplified products were visualized in ethidium bromide-stained 1% agarose gel. The DNA pools that yielded amplified products with fragments of approximately 560 bp were selected.

3.3.1.3 Illumina sequencing and processing

The amplified PCR product was initially cleaned using AMPure XP magnetic beads (Beckman Coulter, Massachusetts, USA) according to the manufacturer's instructions. Following purification, Illumina sequencing dual-index barcodes were added to the amplicon targets using full complement of Nextera XT® indices (Illumina Inc., San Diego, CA, USA) in 25-μL PCR reaction (95 °C, 3 min; 8 x [95 °C, 30s; 55 °C, 30s; 72 °C, 30s]; 72 °C, 5 min; 4 °C, ∞). PCR amplicons were cleaned using AMPure XP® beads, and the fragment size (~630 bp) was validated using the Bioanalyzer DNA 1000® chip (Agilent, Santa Clara, CA, USA) and quantified using the Qubit™ dsDNA HS Assay kit (ThermoFisher Scientific, Waltham, MA, USA). Five microlitre (5 μL) aliquots of purified equimolar DNA (4 nM) from each library were pooled into a single amplicon library and stored at − 20 °C until sequencing. Pooled libraries were subsequently sequenced on the Illumina MiSeq 250 platform with v3 chemistry (2x300 cycles kit) (Illumina Inc., San Diego, CA, USA) at the College of Agriculture and Environmental Sciences (CAES) Eureka Laboratory, University of South Africa, Johannesburg, South Africa.

3.3.1.4. Bioinformatic analyses of Illumina sequencing data

The raw high-throughput sequencing data of the 22 sediment samples analyzed in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database as BioProject ID PRJNA855470 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA855470). Standard techniques based on the Quantitative Insight Into Microbial Ecology (QIIME) v1.7.0 data analysis program (Caporaso *et al*., 2011) were used to conduct the bioinformatics analysis. Briefly, forward and reverse reads were initially merged and truncated to remove the Illumina adaptor and primer sequences. Quality

filtering for each sample, based on average quality score of <20 over a 50 bp sliding window, was used to truncate ~450 bp reads. Reads containing ambiguous characters were removed and quality filtered reads were then compared with the reference database (Ribosomal Database Project (RDP) Gold database) using the USEARCH and UCHIME utility to detect and remove chimeric sequences (Edgar et al., 2011). Chimeric sequences are artefact sequences formed by two or more biological sequences incorrectly joined. The RDP Classifier is a naive Bayesian classifier that can rapidly and accurately provides taxonomic assignments from domain to genus, with confidence estimates for each assignment. The resultant quality-filtered sequence reads were aligned against the SILVA 16S rRNA gene reference database (version 132) with a confidence threshold of 80% (Quast et al., 2013). Finally, clustering of operational taxonomic units (OTUs) at 97% sequence identity was done using the furthest neighbor algorithm. OTU refers to operational definition used to classify closely related sequences (having >97% similarity). Sequences were rarefied prior to calculation of alpha and beta diversity statistics. Abundant OTUs (>0.01 relative abundance) at the genus level were utilized in generating the heat map to visualize dominant taxonomy and determine dissimilarity of bacterial communities between the river catchment samples. For downstream analysis, the sequences were rarefied to lowest sequence read sufficient capture the whole bacterial community. Rarefaction involves random subsampling of sequences from the initial sample library to selected normalized library size.

3.3.1.5 Data visualization and statistical analysis

For visualization of the differences in relative abundance of sediment bacterial community composition, stacked bar charts and heatmap based on the major OTUs at phylum, class, order, family, and genus level were generated by *ggplot2* (Wickham, 2016) and *heatmap.2* packages (Warnes et al. 2019) in R software version 3.6.1 (http://www.r-project.org), respectively, to visualize the variations and distributions of bacterial communities. Alpha diversity of sediment bacterial community was evaluated using rarefied datasets of 18,172 based on observed OTUs, the abundance-based coverage estimator (ACE), Chao1 richness index, Shannon-Weiner (H′) community index, Simpson evenness (D), Phylogenetic Diversity (PD) and Good's coverage index; variation was tested statistically using either one-way ANOVA and post-hoc tests (Tukey-Kramer at 0.95), an effect size (Eta-squared) and multiple test correction using Benjamini-Hochberg FDR (false discovery rate) procedure. As amplicon sequencing data consist of discrete counts of sequence reads, with groups of samples typically having different library sizes that are

not representative of biological variation, library size normalization is required to meaningfully compare diversity between them (Schloss et al., 2009).

For beta diversity analysis, soil bacterial community dissimilarity between the sediment samples was performed by non-metric multidimensional scaling (NMDS) and redundancy analysis (RDA), respectively, based on Bray-Curtis distance matrix using vegan package in R (Oksanen et al., 2010). The significance of the convergence between stream catchment zones physicochemical characteristics and sediment bacterial community was validated by performing nonparametric analysis of similarity (ANOSIM) and adonis PERMANOVA (Anderson and Walsh, 2013) using *vegan* package in R. While ANOSIM generates an *R* statistic and *p* value, adonis PERMANOVA gives both *F* statistic, R^2 and *p* values. Both *R* and R^2 are statistics for compositional dissimilarity, where higher values indicate dissimilarity.

A Venn diagram, generated using *ampVis2* package (Andersen et al., 2018) in R software, was used to evaluate the shared and unique core microbiome at genus level (herein defined as OTUs present in at least 50% of the samples of each group at 1% minimum relative abundance). The possible biomarkers with significant differences were identified using the linear discriminant analysis (LDA) effect size (LEfSe) approach (Segata and Huttenhower, 2011). Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant method used to find a linear combination of features that characterizes or separates two or more classes of objects or events. Linear Discriminant Analysis Effect Size (LEfSe) is an algorithm for high-dimensional biomarker discovery that identifies genomic features, characterizing the differences between two or more biological conditions (Segata and Huttenhower, 2011). First, the algorithm robustly identifies features that are statistically different among biological classes, then performs additional tests to assess whether these differences are consistent with respect to expected biological behavior (Segata et al., 2011). Linear Discriminant Analysis Effect Size uses LDA to estimate the effect size of each differentially abundant feature (Segata et al., 2011).

3.3.2 Determining the sediment microbial communities' functional profiles and the presence of potential bacterial pathogens

3.3.2.1 PICRUSt metagenomic function profiling of bacterial communities

The functional capabilities of identified bacterial communities in sediment samples were also predicted using phylogenetic study of communities by reconstruction of unobserved states

(PICRUSt) (Douglas et al., 2018). Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) uses evolutionary modelling to predict metagenomes from 16S data and a reference genome database (Langille et al., 2013). It uses an extended ancestral-state reconstruction algorithm to predict which gene families are present and then combines gene families to estimate the composite metagenome (Langille et al., 2013). Metagenome predictions were made by corresponding the 16S rDNA marker gene data and the reference genomes in databases, including both Cluster of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. One-way ANOVA and post-hoc tests (Tukey-Kramer at 0.95), an effect size (Eta-squared) and multiple test correction using Benjamini-Hochberg FDR (false discovery rate) procedure was used to detect the differentially abundant KEGG pathways between the threecatchment zones.

3.3.2.2 Presence of potential bacterial pathogens

National Center for Biotechnology Information-basic local alignment search tool (NCBI-BLAST) algorithm was used to search the human bacterial pathogen database (HBPD) for human bacterial pathogens from Illumina sequencing sequences (Cai and Zhang 2013). For the analysis, only sequences with a blast hit of >97 percent identity was taken into account. A bacterial genus was regarded as potentially harmful if at least one species with a minimum abundance of 10 strains classified as biosafety level 2 or 3 by the American Biological Safety Association (https://my.absa.org/tiki-index.php?page=Riskgroups). According to American Biological Safety Association, risk group 2 (RG2) are agents that are associated with human disease which is rarely serious and for which preventive or therapeutic interventions are often available. Risk group 3 (RG3) are agents that are associated with serious human disease for which preventive or therapeutic interventions may be available.

3.3.3 Identifying the key environmental factors accounting for the variations in sediment bacterial community composition and structure

3.3.3.1 Stream water physicochemical parameter analyses

The physical and chemical characteristics of water samples such as pH, electrical conductivity (EC), dissolved oxygen (DO) and turbidity were measured on site using a HANNA

H1991300 multi-parameter ion-specific meter (Hanna Instruments (Pty) Ltd, Bedfordview, South Africa).

3.3.3.2 Environmental factors (sediment physicochemical parameter) analyses

In the laboratory, the wet soil sediment samples were spread on aluminium foil that had been sanitized using 70% ethanol and then dried in the oven at 105°C until a constant weight (Kinuthia *et al*., 2020). Non-soil materials such as stones, plastics, pieces of plant residues including the roots, leaves and stems were manually removed from the sediment samples. Sediments were then grounded using pestle and motor sieved through a 2-mm mesh and subsequently divided into two portions. In each portion, about 50g of sieved sediments were packed in small brown envelopes and labeled.

To measure the pH of the sediment samples, sediment slurry was prepared in deionized water at sediment water ratio 1:5. The slurry was shaken at 150 rpm for 2 hours at room temperature. Once settled (after 12 h), the liquid part was carefully decanted and used for pH analysis using using Adwa AD11 pH meter (Adwa Instruments, Szeged, Hungary). Total organic carbon (TOC) was measured by combusting 10mg aliquots at 680°C in an Apollo 9000 Total Organic Carbon Analyzer equipped with a boat sampler (Model 183 Teledyne Tekmar, Mason, HO). LECO Trumac® Carbon, Nitrogen, and Protein analyzer fitted with boat sampler (Series 828 LECO, Michigan, US) was used for the analysis of total nitrogen (TN) according to manufacturer's instructions.

Inductively coupled optical emission spectrometer (Agilent Technologies 700 Series ICP-OES) analysis of the heavy metal concentration was performed on 0.5g of soil samples after acid microwave digestion as described previously (Sekhohola-Dlamini *et al*., 2020). Briefly, 0.5 g of each sediment sample was digested with concentrated nitic acid (9 mL) and hydrochloric acid (3 mL) in a PTFE microwave bombs in SINEO MDS-6G Microwave (Jinan Hanon Instruments Co., Ltd, Jinan, China) at 175°C for 30 minutes at 6W power as per the guidelines provided by the manufacturers. The digested were filtered and the volume of filtrate quantitavely made up to 50 mL with deionized water. An inductively coupled optical emission spectrometer (Agilent Technologies 700 Series ICP-OES) was used after proper dilutions to measure the heavy metal contents.

3.3.3.3 Heavy metal pollution assessment

Nemerow integrated pollution index (P_N) (Hu et al., 2018) was adopted to evaluate overall heavy metal pollution status in the sediments. P_N comprehensively reflects the status of multiple HMs pollutants and highlights the highest concentration pollutants on the environment. Initially, single factor pollution index (P_i) Pb, Cd, Cr, Cu and Zn was calculated as follows:

$$P_i = \frac{C_i}{S_i} \tag{1}$$

where C_i and S_i are the concentrations of evaluated metal(loid)s in the sample and local background, respectively. The greater the single factor index, the higher level of the contamination of the single heavy metal. The P_i was used to calculate P_N as follows:

$$P_N = \sqrt{\frac{(\frac{1}{n}\sum_{i-1}^{n} P_i)^2 + P_{imax}^2}{2}} \tag{2}$$

where $\frac{1}{n}\sum_{i-1}^{n} P_i$ and P_{imax} are the mean and maximum values of P_i of all the evaluated metal(loid)s, respectively, while n is the number of evaluated pollutants.

Pollution of the sediment by the heavy metals was classified into five grades based on the Nemerow pollution index ($P_N <$ 0.7 — non-pollution; P_N 0.7–1.0 — pollution warning line; P_N 1.0–2.0 — low level of pollution; P_N 2.0–3.0 — moderate level of pollution; $P_N >$ 3.0 — high level of pollution) (Hu et al., 2018).

3.3.3.4 Data visualization and statistical analysis

Statistical Analysis Software (SAS) version 9.4 (SAS Institute, Cary, NC, USA) was used to do a one-way analysis of variance (ANOVA) with Tukey's Honest post hoc test ($p < 0.05$) to compare the physicochemical results. The differences were visualized using histograms created using Prism GraphPad (Vers 8.2.1, GraphPad Software, La Jolla California, USA).

Redundancy analysis (RDA) was used to visualize the stream sediment microbial community structure from the three catchment zones and to determine its correlation with environmental factors such as TP, TN, TOC, pH and heavy metals (Pb, Cd, Cr, Cu and Zn). Redundancy analysis is a constrained method based on multiple linear regressions to extract and summarize the variation in a set of response variables which can be explained by a set of

explanatory variables. Prior to RDA analysis, OTU count data were Hellinger-transformed as implemented in the R package vegan v2.4–1 (Oksanen et al., 2010), and the matrices analyzed using detrended correspondence analysis (DCA) to evaluate the gradient size of the genus distribution. In this study DCA indicated linearly distributed data (length of gradient < 3), suggesting the RDA as the best-fit mathematical model for the data. Forward selection and the Monte Carlo permutation test were applied with 1,000 random permutations to verify the significance of environmental parameters upon the biological variables. Contribution of highly correlating OTUs ($p < 0.01$) with redundancy axes was identified using the envfit functions from the R package vegan. RDA plots were generated using R package vegan v2.4–1 and ggplot2 (Wickham, 2016). To further investigate the correlation between Bray–Curtis dissimilarity distance and environmental factors, linear regression was done using the R package ggpubr (Kassambara, 2020).

CHAPTER FOUR

RESULTS

4.1 Sediment bacterial community structure and diversity along the river/stream catchment using the 16S rRNA gene-based Illumina MiSeq sequencing.

4.1.1 Richness and diversity of sediment microbiome

Illumina sequencing of the 22 samples, delineated into three groups on the basis of urban pollution, resulted in a total of 1,321,460 reads and 896,567 high-quality sequences with an average of 39,719±2,964 sequences and read length of 410±50 bp. The detailed global sequencing data has been provided in Table S1 in the Appendices. Both rarefaction curves and rank abundance plots asymptotically approached a plateau with increase in the number of sequences, suggesting that the curves accurately reflected the whole bacterial communities. Good's coverage across all the samples was >99% (Table 4.1), indicating that the sampling depth was sufficient to estimate the bacterial diversity enclosing all major group inhabiting the three stream catchment systems.
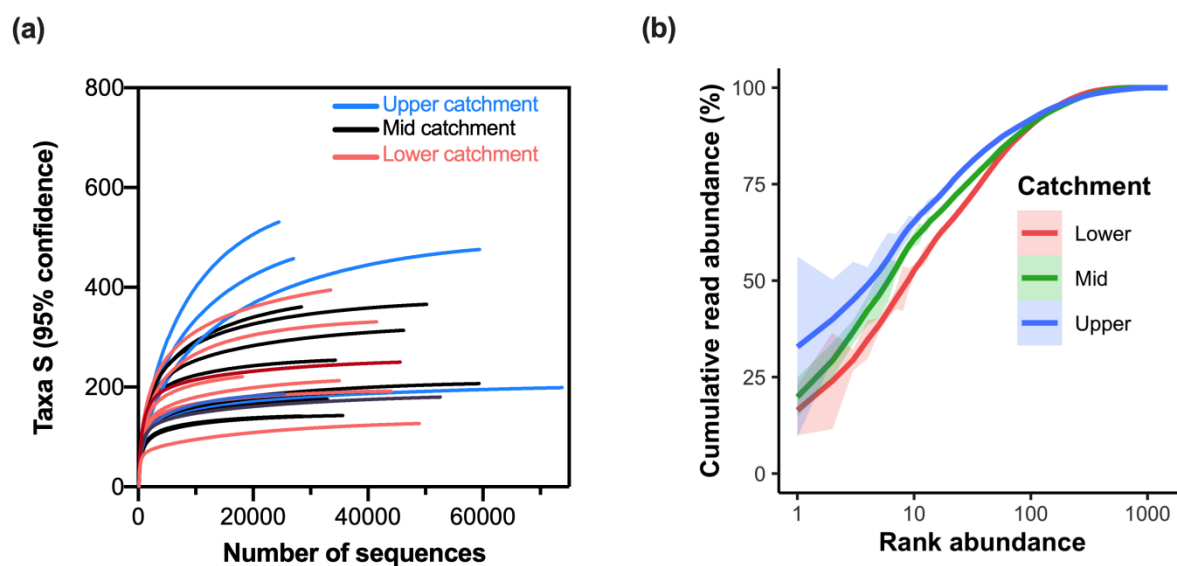
Figure 4.1 Rarefaction curves of operational taxonomic unit (OTUs) diversity of each sample (a) and rank abundance plot of the OTU accumulation for each river catchment (b).

Table 4.1 show the summary of the sequencing data and alpha diversity indices of the sediment bacterial community across the different stream catchment zones. Prior to calculating the

alpha and beta diversity statistics, sequences were rarefied to the lowest sequence read (18,172 sequences). Both species evenness (Simpson) and species richness indices (Observed OTUs, ACE, Chao1 and PD) were significantly higher ($p < 0.5$) for the lower catchment zone stream sediment samples, indicating a richer and diverse bacterial community than Upper and Mid catchment zones. However, no significant differences were observed in Shannon indexes ($p \geq 0.05$) between the different catchment zones of the two streams.

Table 4.1 Summary of sequencing outputs and diversity indices for sediments bacterial communities from different stream catchments flowing through urbanized Kisumu City

Indices	Catchment			p-value
	Lower	Mid	Upper	
Quality reads	46 534±12 126	40 709±2 964	34 499±4 532	
Observed OTUs	1 414±339[a]	704±59[b]	807±116[b]	**
ACE	1 574±350[a]	749±64[b]	883±135[b]	**
Chao1	1 492±339[a]	721±61[b]	842±127[b]	**
Simpson (D)	0.15±0.05[a]	0.04±0.01[b]	0.04±0.01[b]	**
Shannon (H)	3.92±0.25	4.32±0.10	4.50±0.12	
PD	1 563±265[a]	864±87[b]	1 055±146[ab]	*
Good's coverage (%)	99.7	99.2	99.4	

[†] *Sample comparison for diversity indices (observed OTUs, ACE, Chao1, Shannon, Simpson evenness and Phylogenetic Diversity) were performed using rarefied datasets of 18,172 sequences representing the lowest number of reads in a sample. One-factor ANOVA was carried out to compare the sampling sites and post hoc (Tukey's) tests applied. Different superscript letters within a row indicate significant differences at $p \leq 0.05$, * ; $p \leq 0.01$, **; and $p \leq 0.001$, ***. PD, phylogenetic diversity.*

4.1.2 Bacterial compositional structure across the catchment zones

A total of 36 phyla, 91 classes, 193 orders, 429 families and 1483 genera were identified in the sediment bacterial community across the different stream catchments. The stream sediment bacterial community composition was dominated by the phylum *Actinobacteria* (3.4-66.5%), *Proteobacteria* (9.4-75.0%), and *Firmicutes* (3.8-50.6%), followed by *Bacteriodetes* (0.1-16.4%), *Cyanobacteria* (0.0-20.0%), *Verrucomicrobia* (0.0-10.3%), *Fusobacteria* (0.0-1.5%), *Acidobacteria* (0.0-1.3%), *Chloroflexi* (0.0-1.1%) and *Gemmatimonadetes* (0.0-0.5% (Figure

35

4.2a). The relative abundance of all phylum detected in the sediment samples is provided in Table S2 in the Appendices. The major classes detected in the sediment samples included unclassified *Actinobacteria*, *Bacilli*, *Alphaproteobacteria, Gammaproteobacteria, Clostridia, Sphingobacteria* and *Bacteroidia* (Figure 4.2b).
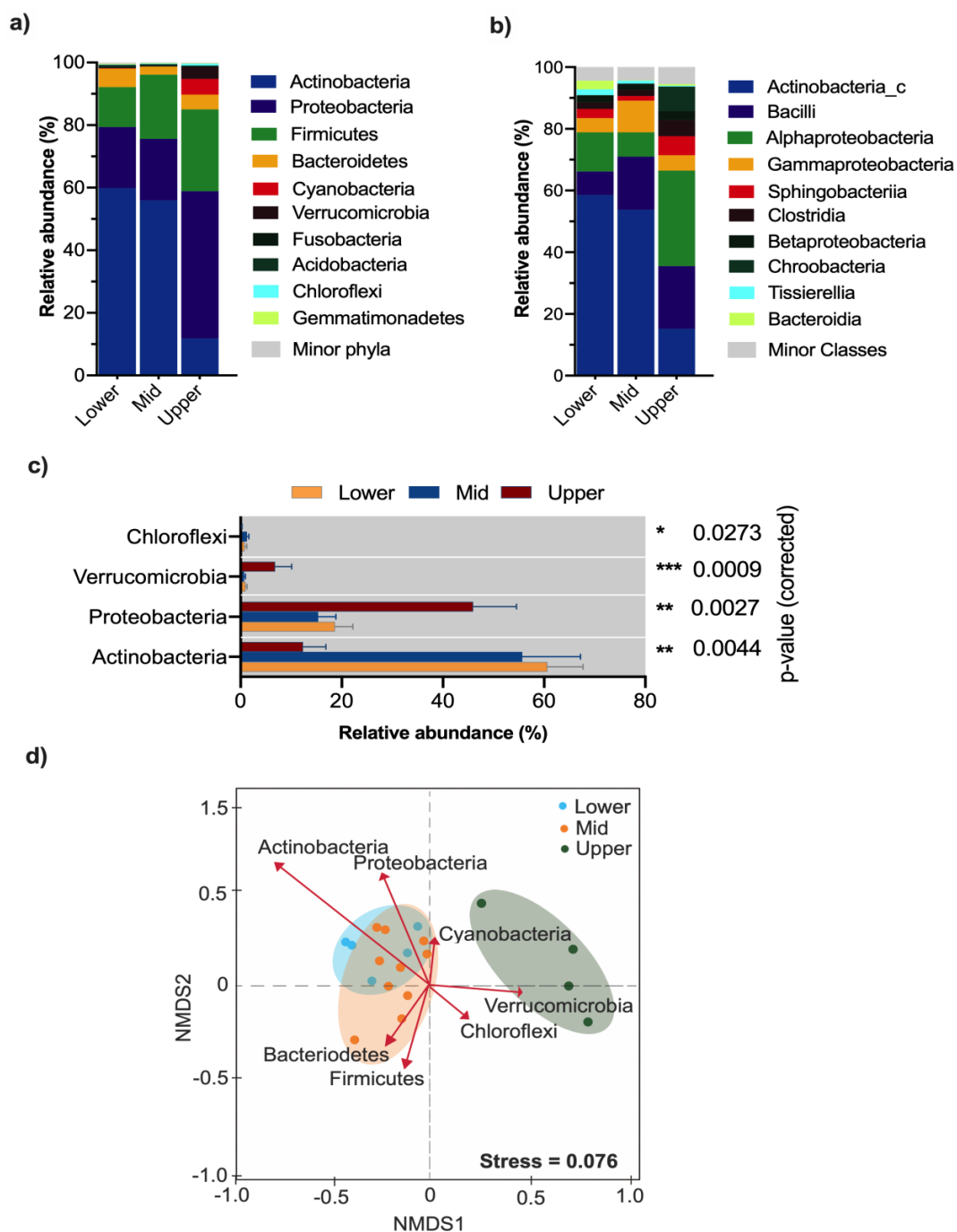
Figure 4.2 Composition and diversity of river sediment bacterial community. a) Relative abundance at phylum level and class level (b). c) Significantly different relative abundance of dominant bacterial phyla (having > 0.1% relative abundance) in three groups of stream catchment sediments. One-way ANOVA test and post-hoc Tukey's test with subsequent Benjamini-Hochberg false discovery rate (FDR) correction was used to evaluate the importance of comparisons between indicated groups. *, $p < 0.05$; **, $p < 0.01$; and ***, $p < 0.001$. c) The non-metric multidimensional scaling (NMDS) plot based on Bray–Curtis dissimilarity showing significant differences in similarity tested by ANOSIM ($R = 0.5085$, $p < 0.0001$) and adonis PERMANOVA ($F = 4.77$, $R^2 = 0.334$, $p = 0.001$).

Differences in sediment bacterial phyla abundances was observed in the upper, mid, and lower catchment zones of the two streams (Figure 4.2b), suggesting that bacterial diversity was differentially affected by stream catchment conditions. The mid and lower catchment zone sediment of the two streams had higher enrichment of phylum *Actinobacteria.* In contrast, the upper catchment zones had significant enrichment of phylum *Proteobacteria* and *Verrucomicrobia,* while members of phylum *Chloroflexi* were significantly ($p < 0.5$) enriched in the mid catchment zone. For a further glimpse into the compositional and structural differences of sediment bacterial communities in the catchment zones, beta diversity analysis based on NMDS was employed. To test the significance of the observed spatial heterogeneity, NMDS (based on Bray-Curtis dissimilarity), ANOSIM, and *adonis* PERMANOVA analysis of community abundance at phylum level was performed (Figure 4.2c). Overall, NMDS plots exhibited a clear separation of sediment samples in the ordination space according to the catchment zones than the stream. This was further supported by analysis of similarity (ANOSIM) and adonis PERMANOVA which indicated that sampling sites differed significantly in dispersion when community structure was considered (ANOSIM $R = 0.5085$, $p < 0.0001$: PERMANOVA, $F = 4.77$, $R^2 = 0.334$, $p = 0.001$).

The sediment bacterial composition at the order and family level has been provided in Figure 4.3. At the family level, *Corynebacteriaceae, Micrococcaceae, Staphylococcaceae* and *Propionibacteriaceae* were the most dominant taxa in the lower and mid catchment stream sediments accounting for > 35% of recovered sequences (Figure 4.3b). In contrast, the upper catchment stream samples microbiome was dominated by Proteobacterial (*Methylocystaceae*

(33%), *Xanthomonadaceae* (4.7%)), *Firmicutes* (*Bacillaceae* (13.3%) and *Clostridiaceae* (4.5%)) and Bacteriodetes families (*Sphingobacteriaceae* (3.9%)).

(a)

	Lower	Mid	Upper
Actinobacteria; Corynebacteriales	19.6	23.8	3.9
Actinobacteria; Micrococcales	16.2	16.9	2.9
Proteobacteria; Rhizobiales	8.5	4.6	36.8
Firmicutes; Bacillales	5.5	11.2	17.6
Actinobacteria; Propionibacteriales	8.1	8.9	1.5
Firmicutes; Lactobacillales	3	5.7	0.9
Actinobacteria; Streptomycetales	4.3	3.3	1.3
Proteobacteria; Pseudomonadales	1.8	4.3	0.7
Actinobacteria; Pseudonocardiales	8.4	1.1	0.9
Firmicutes; Clostridiales	2	1.7	5.9
Bacteroidetes; Sphingobacteriales	2.7	1.5	3.9
Proteobacteria; Xanthomonadales	1.2	1	4.7
Proteobacteria; Enterobacterales	0.6	2.2	0.1
Proteobacteria; Sphingomonadales	1.9	1.5	0.2
Proteobacteria; Burkholderiales	1.3	0.7	2
Firmicutes; Tissierellales	1.9	0.9	0.4
Proteobacteria; Neisseriales	1.2	1.2	0.3
Proteobacteria; Pasteurellales	0.6	1.4	0.1
Cyanobacteria; Oscillatoriales	0	0	5
Bacteroidetes; Bacteroidales	2.9	0.1	0.5

(b)

	Lower	Mid	Upper
Actinobacteria; Corynebacteriaceae	18.8	23.1	3.7
Actinobacteria; Micrococcaceae	7.2	8.6	1.1
Firmicutes; Staphylococcaceae	3.9	9.5	0.9
Proteobacteria; Methylocystaceae	0	0	33
Actinobacteria; Propionibacteriaceae	5.6	7.7	0.9
Firmicutes; Bacillaceae	1.1	1.2	13.3
Actinobacteria; Streptomycetaceae	4.3	3.3	1.3
Proteobacteria; Moraxellaceae	1.7	4.1	0.7
Firmicutes; Streptococcaceae	1.6	4.1	0.7
Actinobacteria; Brevibacteriaceae	3.2	3.4	0.4
Actinobacteria; Pseudonocardiaceae	8.4	1.1	0.9
Actinobacteria; Microbacteriaceae	3.2	2.9	0.6
Bacteroidetes; Sphingobacteriaceae	2.2	1.1	3.9
Proteobacteria; Xanthomonadaceae	1.2	1	4.7
Actinobacteria; Nocardioidaceae	2.5	1.3	0.6
Proteobacteria; Rhizobiaceae	2.1	1.2	1.6
Firmicutes; Clostridiaceae	0.6	0.2	4.5
Firmicutes; Peptoniphilaceae	1.9	0.9	0.4
Proteobacteria; Neisseriaceae	1.2	1.2	0.3
Firmicutes; Aerococcaceae	1.1	1.2	0.1

Figure 4.3 Relative abundance of bacterial community at order (a) and family level (b)

The relative abundance of the bacterial communities at the genera level (Figure 4.4a) also provided more insight into the taxonomic differences between the different stream catchments. Overall, the top 10 genera across all the stream catchments were in the order: *Corynebacterium* > *Staphylococcus* > *Methylocystis* > *Cutibacterium* > *Auritidibacter* > *Brevibacterium* > *Acinetobacter* > *Streptococcus* > *Streptomyces* > *Turicella*. A dendogram heatmap plot showing complete-linkage agglomerative clustering based on a Euclidean distance of the top 50 genera is illustrated in Figure 4.4a. Comparatively, genus *Corynebacterium* was the most abundant taxa in lower and mid catchment, constituting ~38.6% of the total sequences. Other major genera identified in stream samples included *Staphylococcus* (14.1%), *Methylocystis* (32.8%), *Cutibacterium* (13.6%) and *Auritidibacter* (11.8%). In addition, both lower and mid catchment had higher abundance of *Cutibacterium* (5.4% *vs.* 7.4%) and *Auritidibacter* (5.3% *vs.* 5.6%). Enrichment of genus *Amycolatopsis* (7.6%) was also detected in lower catchment zones.

Interestingly, higher abundance of genus *Methylocystis* (32.8%) and *Bacilli* (7.1%) was detected in the upper stream catchment.
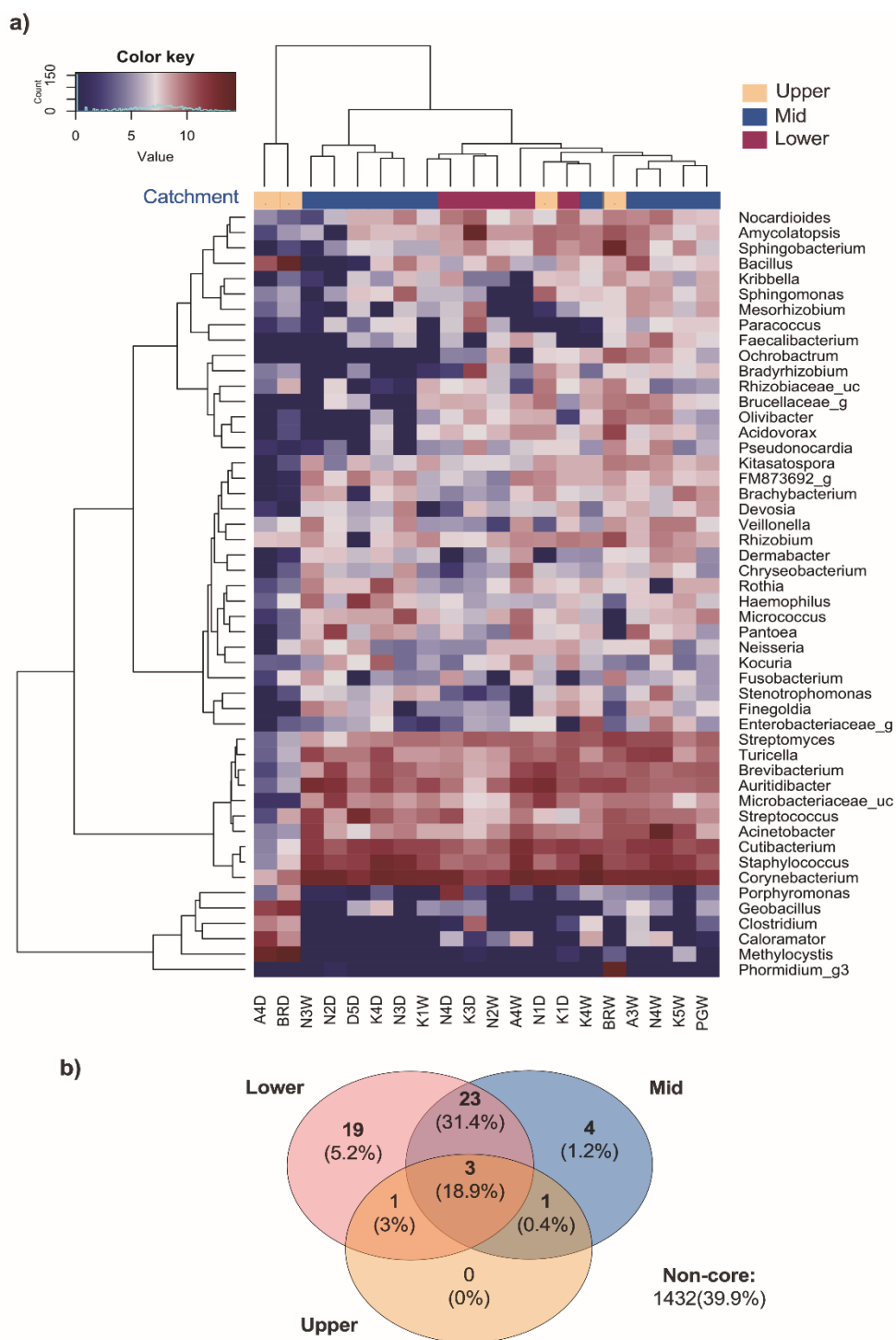
Figure 4.4 Relative abundance of bacterial community at order (a) and family level (b). a) Heatmap of log$_2$ normalized counts of the 50 most abundant genera in the sediments of three stream

catchments. The heatmap color (blue to reddish-brown) represents the row z-score of the mean relative abundance from low to high. **b)** Venn diagram showing the unique and shared core microbiome.

4.1.3 Core microbiome of stream sediments

As illustrated in Figure 4.4a, hierarchical clustering of the top 50 genera revealed no clear clustering of samples on the basis of catchment zones. This result suggests a shared or co-occurrence of sediment bacterial genera between the three catchment zones. To support this finding, a Venn diagram was used to evaluate the shared and unique sediment core microbiome (Figure 4.4b). A total of 46, 31 and 5 OTUs were obtained as core microbiome for the lower, mid and upper catchment, respectively, representing 60.1% of the total reads generated for sediment bacterial community. Overall, there were 3 core OTUs (unclassified *Corynebacteriaceae*, *Rhizomicrobium* and *Staphylococcus*) accounting for 18.9% of the total OTUs that were shared among the three catchments. However, differences in variations in the unique OTUs were observed between the catchment zones. Curiously, no unique core OTUs were detected in the upper nonpolluted stream catchment sediments. In contrast, mid and lower catchments had 4 (1.2%) and 19 (5.2%) unique OTUs, respectively. In contrast, mid and lower catchments had 4 (1.2%) and 19 (5.2%) unique OTUs, respectively. The bulk of these taxa included *Rothia, Stenotrophomonas, Granulicatella* and *Dermabacter Acidovorax, Phycicoccus, Pseudolabrys, Aeromonas, unclassified Rhizobiaceae, Shinella, Variibacter, Bradyrhizobium, unclassified Brucellaceae, Enhydobacter, Kocuria, Luteimonas, Microbacterium, Mycobacterium, Nakamurella, Olivibacter* and *Facklamia.*

4.1.4 Taxonomic biomarkers of stream sediments

Biomarker analysis with LEfSe analysis was performed to identify the OTUs with significant abundance differences in the different types of catchment sediments samples of the two streams. Overall, 31 differentially abundant taxa (LDA > 2.0, $q < 0.01$) were detected, including 8 orders, 10 families and 12 genera level biomarkers across the catchment zones (Figure 4.5a). The Cladogram showing the metagenomic biomarkers at phylum, class, order, family and genus levels is illustrated in Figure 4.5b.

(a)
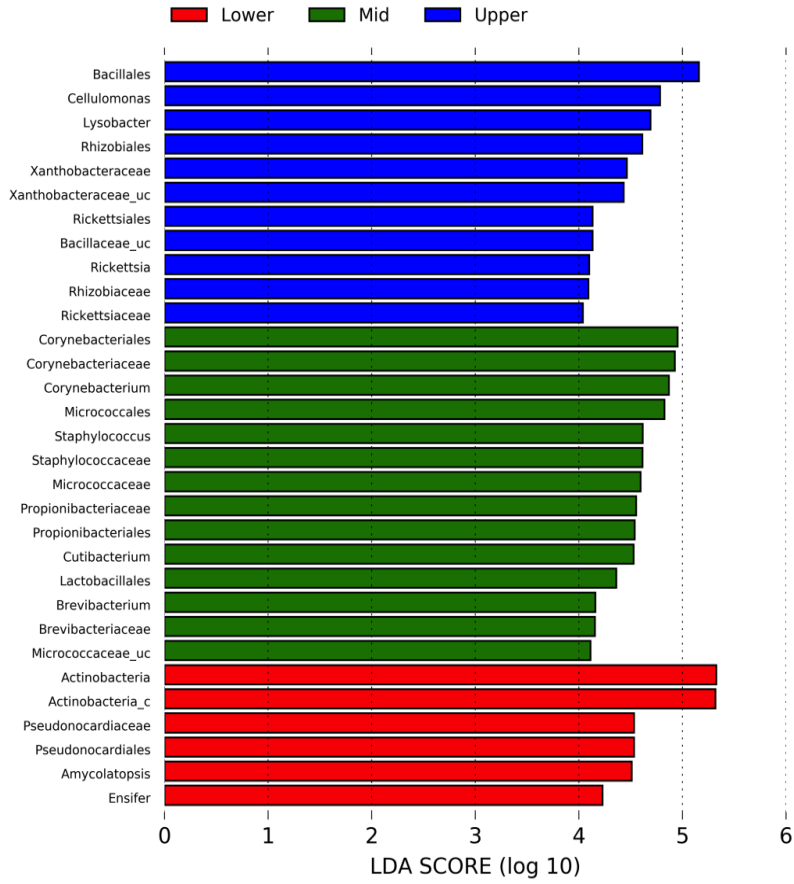
(b)
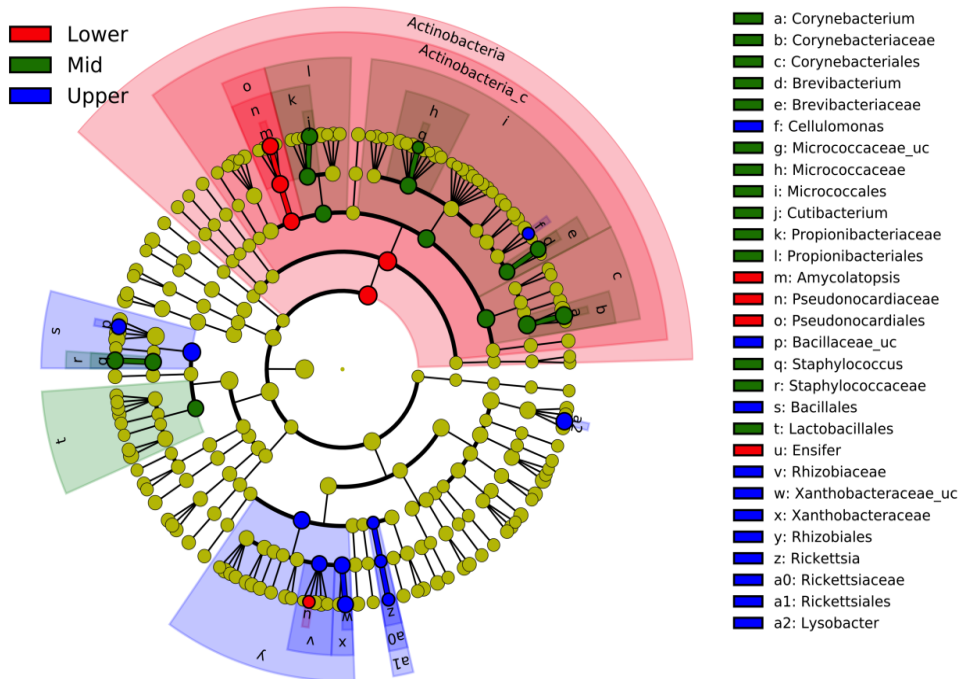
Figure 4.5 Linear Discriminant Analysis (LDA) effect plot and cladogram showing differential and phylogenetic lineages of indicator bacteria associated with sediments from three river catchments. (a) Indicator bacteria with LDA effect size ≥ 2 associated with sediments from three river catchment zones. **(b)** Cladogram showing the phylogenetic distribution of the bacterial lineages in river sediments under different catchments. The phylum, class, order, family, and genus levels are listed in order from the inside to the outside of the cladogram. Different-colored nodes correspond to different sample groups, which represent taxa with significant enrichment in the corresponding group and significant influence on intergroup difference (LDA effect size > 2.0). Yellow circles stand for taxa with no significant differences in the sediment.

Overall, upper and mid catchment zones had the largest number of taxonomic biomarkers. Genus *Cellulomonas, Lysobacter, Rickettsia, unclassified Xanthobacteriaeceae and unclassified Bacillaceae* were the key taxonomic biomarkers in the upper zone. In contrast, only genus *Ensifer* and *Amycolatopsis* were the main taxonomic biomarkers in lower catchment zone.

4.2 Sediment microbial communities' functional profiles and the presence of potential bacterial pathogens in the different stream catchment zones

4.2.1 PICRUSt metagenomic function profiling of sediment bacterial communities

The biological function of sediment bacteria related to metabolism, genetic information processing, environmental information processing, cell processes, human diseases, and organism systems was predicted using metagenomic profiling based on the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States Version 2 (PICRUSt2) algorithm (Douglas *et al.*, 2018). The most abundant level 2 Kyoto Encyclopedia Genes and Genomes (KEGG) pathways related to metabolism (carbon, sulfur, nitrogen and methanol) and disease pathogenesis detected are shown in Figure 4.6a. In addition, the three stream catchment sediments had comparable abundance of pathways related to metabolism (carbon, sulfur, nitrogen and methanol) and disease pathogenesis. However, significant differences ($p < 0.05$) in the abundance of key pathways were identified, mainly related to nitrogen metabolism and pathogenesis. Whereas, disease/virulence markers such as polymyxin resistance, peptidoglycan biosynthesis V (beta-lactam resistance) and II (staphylococci), and nitrifier denitrification pathway were significantly

enriched in lower and mid catchment zone, upper zone sediment was characterized by higher assimilatory nitrate reduction I and urea cycle pathways.
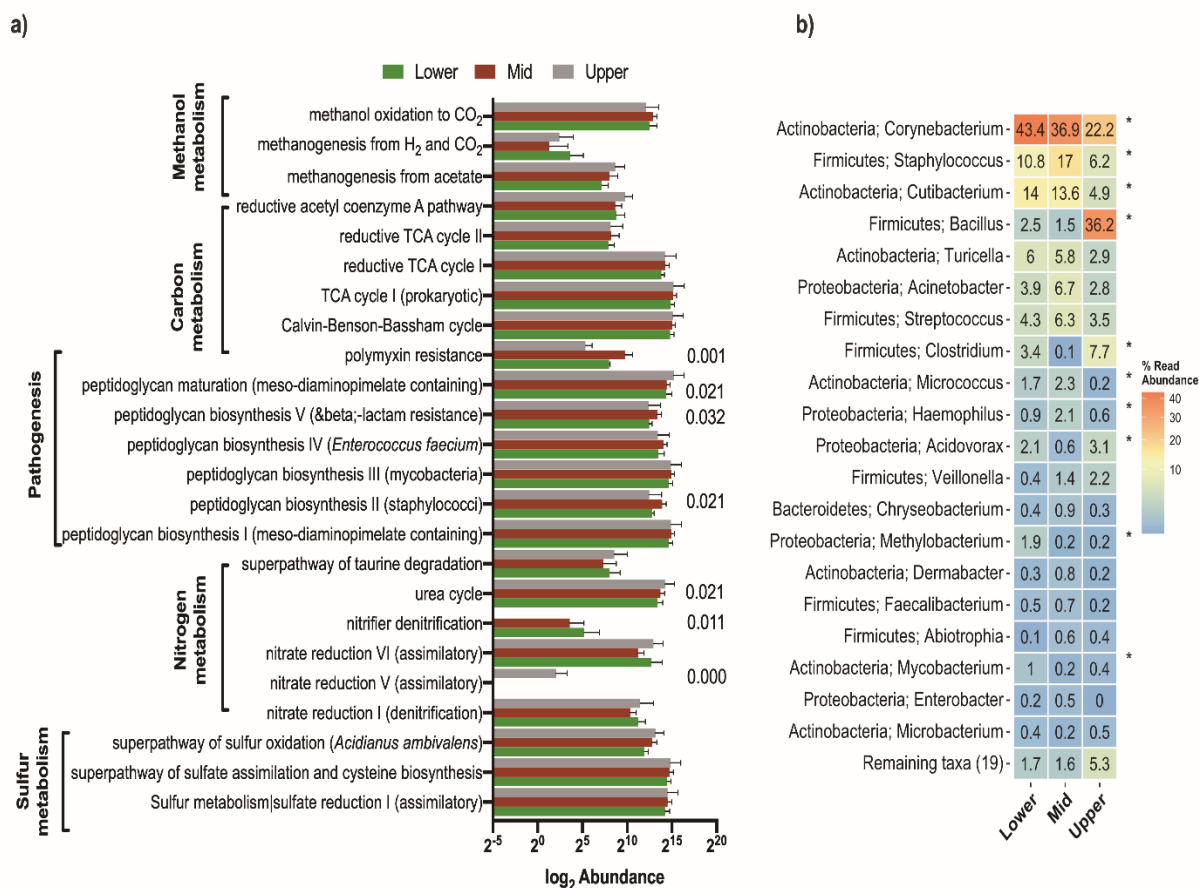
Figure 4.6 Functional prediction and potential pathogens distribution in the stream sediments. a) The level 2 KEGG metabolic pathways of bacterial communities in different stream catchment sediments predicted by PICRUSt2. *P*-values denoting the significant difference in metabolic pathways between the catchment sediment are provided. **b)** Heatmap showing the relative abundance of major potential pathogens (>0.1 % average abundance ratio) in the different catchment sediment samples. The asterisk indicated significant difference in abundance between different catchments ($p < 0.05$).

4.2.2 Potential bacterial pathogens distribution

In this study, a total of 39 potential pathogenic bacterial genera were identified (Figure 4.6b). The analysis also indicated that 19 of the 39 potential pathogenic genera displayed average abundance ratios of >0.1% in at least one of the catchment zones. Among these taxa, 10 potential

bacterial genera exhibited significant differential enrichment between the different catchment zones. Overall, lower and mid catchment sediments exhibited enrichment of higher number of potential pathogenic groups such as *Corynebacterium*, *Staphylococcus*, *Cutibacterium*, *Turicella*, *Acinetobacter*, *Micrococcus* and *Faecalibacterium*. Compared with other catchment, upper zone had significant enrichment ($p < 0.05$) of *Bacillus*, *Clostridium* and *Acidovorax*, whereas *Methylobacterium* and *Mycobacterium* exhibited significant enrichment ($p < 0.05$) in the lower zone. Other important potential pathogenic bacteria detected in lower and mid catchment zones, albeit at low abundance, included the key gut-associated bacterial genera such as *Shewanella*, *Escherichia*, *Klebsiella*, *Enterococcus*, *Prevotella*, *Legionella*, *Vibrio*, *Salmonella* and urban wastewater-rich genera such as *Campylobacter, Roseomonas, Burkholderia and Actetobacterium.*

4.3 Environmental factors affecting the variations in sediment bacterial community composition and structure

4.3.1 Physicochemical characteristics of stream water and sediments

The physicochemical parameters of surface water such as conductivity was significantly higher ($p < 0.05$) for the lower and mid than upper catchment zone. On the other hand, turbidity was 1.5- and 10.9-fold higher in lower catchment, than mid and upper catchment, respectively. Similarly, ammonium nitrogen (NH_4^+-N) levels were 1.8- and 2.8-fold higher in the lower catchment, than in the mid and upper catchment zone, respectively. In the lower catchment sampling sites, a characteristic intense (foul) smell was noted, indicating the gross contamination of the streams catchment with ammonium nitrogen rich wastes. On the other hand, a significantly higher total organic carbon (TOC), total nitrogen (TN) and total phosphorus (TP) was observed in stream water from the mid and lower catchments zones. In contrast, dissolved oxygen (DO) and pH were significantly higher ($p = 0.0002$) for the upper catchment, than the other two stream catchment zones (Table 4.2). The significant depletion ($p < 0.05$) of DO for the lower catchment zone imply high consumption of available O_2 by the microbial activity and organic pollutants. Only alkalinity levels revealed no significant differences across the different stream catchment zones.

In tandem with the water results, lower and mid catchment zone sediment of the two streams recorded significant lower ($p < 0.05$) pH values and elevated ($p < 0.01$) levels of TOC, TN and TP than upper catchment zone. The highest mean values for Fe, Mn, Zn, Pb, Cu, Cd and

Cr were recorded in mid and lower catchment zones (Table 4.2). However, the highly urbanized lower catchment zone had significant higher levels of Cd, Zn and Cu ($p < 0.05$), but had comparable levels of Pb, Cr and Mn with moderately urbanized mid catchment zones. Comparatively, peri-urban upper catchment sediments samples exhibited two- to ten-fold lower heavy metal contents than other sampling sites, indicating that river water pollution is mainly linked to urban anthropogenic inputs.

Table 4.2 The mean (±SD) Nemerow pollution index for single factor (Pi) and integrated pollution index (PN) for heavy metals in the different stream sediments.

| Parameter [†] | Catchment | | | p-value |
	Lower	Mid	Upper	
Sediments				
Cd (mg/kg, dry mass)	$1.14 \pm 0.07a$	$0.71 \pm 0.07b$	$0.11 \pm 0.03c$	<0.0001
Pb (mg/kg, dry mass)	$126 \pm 15.2a$	$120 \pm 9.10a$	$21 \pm 3.70b$	0.0012
Zn (mg/kg, dry mass)	$152 \pm 15.6a$	$97 \pm 3.54b$	$78 \pm 10.0b$	<0.0001
Cr (mg/kg, dry mass)	$35.4 \pm 5.93a$	$24.6 \pm 3.52a$	$0.03 \pm 0.01b$	0.001
Fe (mg/kg, dry mass)	$1327 \pm 89a$	$1074 \pm 63b$	$1241 \pm 25ab$	0.020
Cu (mg/kg, dry mass)	$140 \pm 36a$	$66 \pm 5.18b$	$26 \pm 3.41b$	0.0011
Mn (mg/kg, dry mass)	$1114 \pm 153a$	$901 \pm 73a$	$482 \pm 18b$	0.0065
TOC (mg/g, dry mass)	$11.8 \pm 1.12a$	$9.82 \pm 1.02a$	$5.45 \pm 1.38b$	0.0002
TN (mg/g, dry mass)	$9.83 \pm 1.24a$	$7.37 \pm 0.63a$	$3.11 \pm 1.0b$	0.0012
pH	$6.98 \pm 0.09a$	$7.01 \pm 0.02a$	$7.33 \pm 0.12b$	0.009
C/N ratio	1.34 ± 0.31	1.47 ± 0.16	2.04 ± 0.49	0.290
TP (mg/g, dry mass)	$2.80 \pm 0.28a$	$1.55 \pm 0.14b$	$0.85 \pm 0.12c$	<0.0001
Surface water				
pH	$7.02 \pm 0.06a$	$7.05 \pm 0.06a$	$7.43 \pm 0.09b$	0.0121
Conductivity (µS/cm)	$831 \pm 76a$	$584 \pm 80ab$	$346 \pm 63b$	0.0286
Alkalinity (mg/L)	443 ± 26	459 ± 29	431 ± 45	0.862
Turbidity (NTU)	$185 \pm 24a$	$120 \pm 16a$	$17 \pm 2.85b$	0.0004
NH_4^+-N (mg/L)	$2.50 \pm 0.41a$	$1.43 \pm 0.10b$	$0.90 \pm 0.18b$	0.0006
DO (mg/L)	$1.56 \pm 0.94a$	$5.84 \pm 0.54a$	$7.70 \pm 0.28b$	0.0002

[†] *One-factor ANOVA was carried out to compare the sampling sites and post hoc (Tukey's) tests applied. Different superscript letters within a column indicate significant differences at $p \leq 0.05$. TOC – total organic carbon; TN – total nitrogen; NH_4^+-N – ammonia nitrogen; and DO – dissolved oxygen. NTU – Nephelometric turbidity units*

[τ] *Consensus-based threshold effect concentration (Simpson et al., 2007)*

4.3.2 Assessment of pollution of the river sediments

According to the pollution assessment standard proposed by Nemerow (Nemerow, 1974), the average of the single factor pollution index (P_i) of five heavy metals in sediments were in the order: Pb > Cd > Cu > Zn > Cr (Figure 4.7). Overall, the P_i value of Pb, Cd and Cu were significantly elevated ($p < 0.05$) mainly in the urbanized lower and mid zones than the peri-urban upper catchment stream sediments. However, "severe pollution" ($P_i > 3$) by Pb, Cd and Cu was recorded in the lower catchment zone. Both mid and lower catchment zones recorded "severe pollution" ($P_i > 3$) by Pb. In contrast, the mid catchment zone recorded "moderate pollution" by Cd ($2 < P_i \leq 3$) and "light pollution" by Zn ($1 < P_i \leq 2$). For the upper catchment, "light pollution" was only recorded for Cd. The integrated pollution index (P_N) values can be used as an indicator of the comprehensive pollution by multiple heavy metals. The average P_N of sediments for lower, mid and upper catchment zones were 4.97±0.98, 4.59±0.47 and 0.84±0.31, respectively, indicating that the lower and mid catchment zones were "severely polluted" ($P_N > 3.0$) with heavy metals (Figure 4.7).
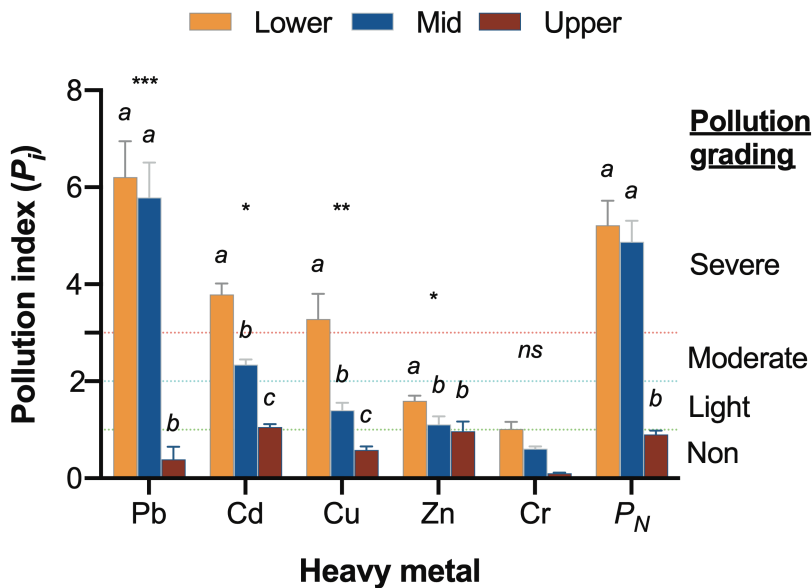
Figure 4.7 The mean (±SD) Nemerow pollution index for single factor (P_i) and integrated pollution index (P_N) for heavy metals in the different stream sediments. One-factor ANOVA

was carried out to compare the P_i of each metal and P_N and post hoc (Tukey's) tests applied. Different italicized letters within a group indicate significant differences at *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; and ns, not significant. Pollution grading based on P_i values is provided.

4.3.3 Environmental factors influencing sediment microbial community structure

Both redundancy analysis (RDA) and regression analysis were used to assess the relationship between stream environmental conditions and the composition of the sediment bacterial population (Figure 4.8). RDA plot revealed that the overall structures of microbial communities in different stream catchments were significantly linked to the sediment properties (Figure 4.8a), with the first two axes (RDA1 and RDA2) explaining 80.82% of the total variance.

RDA ordination analysis resulted into two distinct groups where the polluted lower and mid catchment zones grouped into a cluster, and the upper catchment samples forming the other cluster. The main factors that grouped the lower and mid catchment samples were TOC, Pb, Cd, TN and Cr (Figure 4.8a), key indicators of urbanized anthropogenic pollution. Meanwhile, cluster of upper catchment samples appeared to be mostly influenced by pH. Pb content was the most influential sediment property ($p < 0.01$), followed by TOC ($p < 0.01$) and Cd content ($p < 0.01$). In support of this findings, linear regression analysis also revealed significant moderate positive correlation ($R > 0.4$, $p < 0.01$) between Pb and TOC content with microbiome Bray-Curtis distances (Figure 4.8b).
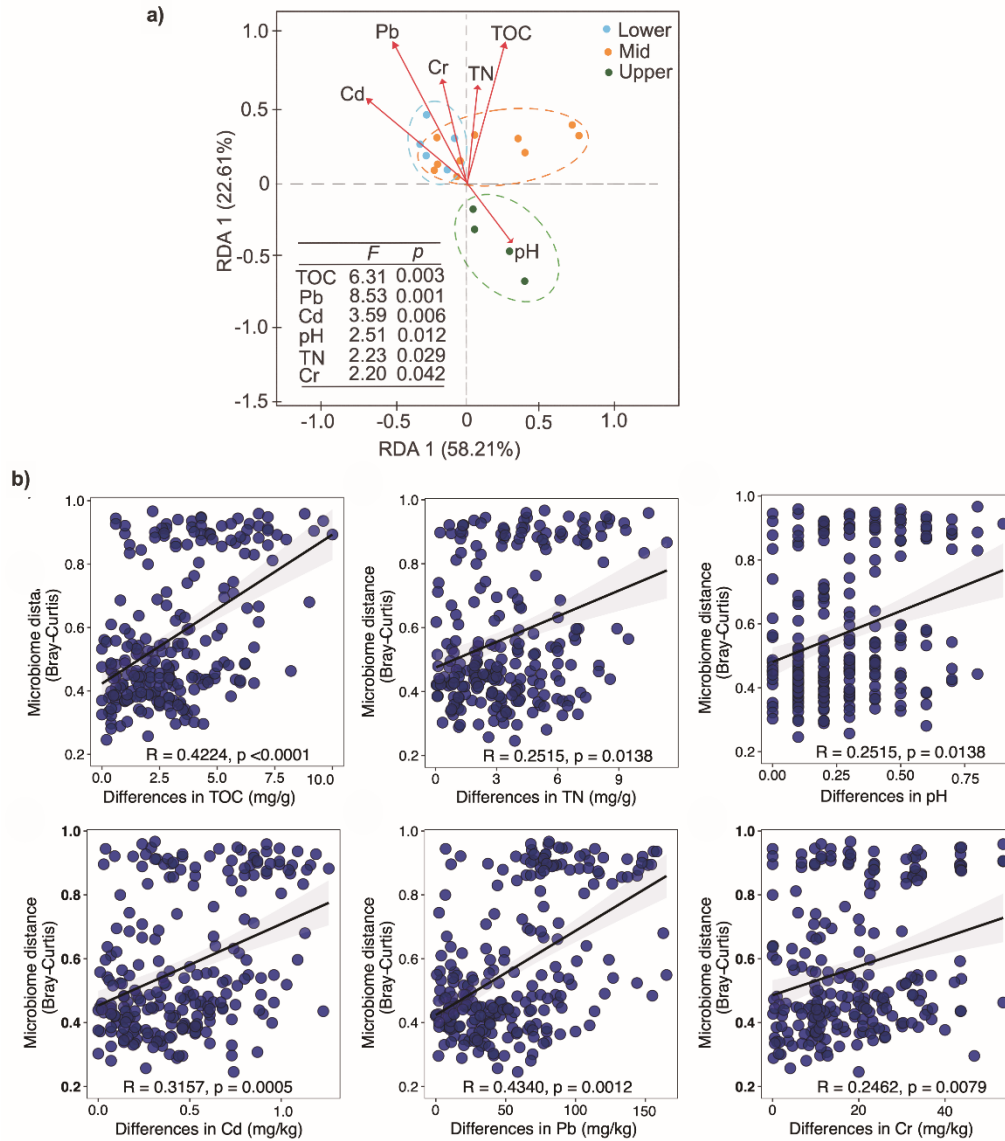
Figure 4.8 Diversity of sediment bacterial community as impacted by selective environmental variables. (a) RDA plot of microbial patterns and six selective variables from samples of lower, mid and upper stream catchment sediments. **(b)** Linear regression of selective environmental variables based on microbiome Bray-Curtis distances.

CHAPTER FIVE

DISCUSSION

5.1 Stream sediment bacterial compositional structure across the catchment zones

5.1.1 Diversity and composition of sediment bacterial community

Sub-Saharan African cities lack a comprehensive understanding of the precise effects of microbial and chemical contamination from anthropogenic input on the composition and organization of sediment microbial communities in urban river ecosystems. In this study, Illumina targeted 16S rDNA amplicon sequencing was used to determine the bacterial composition of two urban streams flowing through Kisumu city into Lake Victoria. The sequencing results showed that the sampling depth was adequate to assess the microbiological diversity containing all main groups inhabiting the three stream catchment systems, as evidenced by Good's coverage over the sample being >99 percent (Table 4.1). Rank abundance plots and rarefaction curves, asymptotically approached a plateau suggesting that sampling size accurately reflected the entire bacterial populations (Figure 4.1).

Alpha diversity indices such as observed OTUs (1414 ± 339), ACE (1574 ± 350), Chao1 (1492 ± 339), Simpson (D) (0.15 ± 0.05) and PD (1563 ± 265) were significantly higher in lower river catchment than mid and upper catchments (Table 4.1). One-factor ANOVA and post hoc (Tukey's) tests results indicated significant differences ($p \leq 0.01$) in bacterial richness (Observed OTUs, ACE and Chao1) and species evenness (Simpson and PD). Numerous studies have demonstrated that urbanization probably increases habitat variability and reduces bacterial diversity (Wang et al., 2017; Zhang et al., 2020). However, other studies have reported that increasing environmental heterogeneity may contribute to higher alpha diversity because it affects the establishment of diffused species and thus reduces the similarity of bacterial community makeup (Tripathi *et al.*, 2018; Yi *et al.*, 2022). Consistent with the latter school of thought, findings of this study, therefore, suggests that urbanization anthropogenic inputs increase the river environmental heterogeneity, that may lead to higher alpha diversity indices in the nutrient-rich urban polluted environments (Jacquiod *et al.*, 2018; Yi *et al.*, 2022).

It is important to know how varying levels of watershed disturbances affect the microbial communities in surface water and sediment since they are crucial for driving the biogeochemical cycles of littoral ecosystems. There is substantial evidence that urbanization affects the variety and

composition of riverine microbes either through dispersal or by altering the environment at the local or regional level (Godoy *et al*., 2020; Hosen *et al*., 2017; Medeiros *et al*., 2016; Numberger *et al*., 2021). The stream sediment bacterial community composition was dominated by the phylum *Actinobacteria, Proteobacteria, Firmicutes, Bacteriodetes, Cyanobacteria*, and *Verrucomicrobia* (Figure 4.2a). The major classes detected in the sediment samples included unclassified *Actinobacteria, Bacilli, Alphaproteobacteria, Gammaproteobacteria, Clostridia, Sphingobacteria* and *Bacteroidia* (Figure 4.2b). A similar taxonomic profile have been reported elsewhere in river sediments impacted by urbanization (Godoy *et al*., 2020; Marcial Gomes *et al*., 2008; Nguyen, 2017). However, subtle variations in the dominant bacterial phyla between the different stream catchment sediments were observed. Bacterial abundances significantly varied between upper, mid and lower catchments ($p < 0.05$, one-way ANOVA test) for the following 4 phyla: *Proteobacteria, Actinobacteria, Chloroflexi* and *Verrucomicrobia,* (Figure 4.2c), suggesting that bacterial diversity was differentially affected by stream catchment conditions. NMDS analyses (based on Bray-Curtis dissimilarity) also revealed clear clustering of sediments samples according to catchment zones in the ordination space (Figure 4d). Specifically, samples from highly urbanized catchment (lower and mid) zones grouping in one cluster, and upper catchment samples into the other cluster. Analysis of similarity (ANOSIM) and adonis PERMANOVA results also that showed substantial differences in dispersion between sampling sites when community structure was taken into account (PERMANOVA, $F = 4.77$, $R^2 = 0.334$, $p = 0.001$). Collectively, the observed spatial differentiations of the samples imply that the stream catchment conditions may be the important determinants of bacterial community composition and structure.

In this study, mid and lower stream catchments of both streams experienced high discharges of wastewater that may originate from household grey water, sewage treatment plants, garages, car washes and industrial waste runoff. *Actinobacteria* was the most abundant phylum in both the lower and mid catchments and was mainly composed of *Corynebacterium*, *Cutibacterium*, *Auritidibacter* and *Brevibacterium* (Figure 4.4a). It has been established that *Actinobacteria* are essential to the carbon cycle in freshwater habitats (Mikhailov et al., 2019). Nguyen (2017) also have reported that actinobacterial species may be crucial in the breakdown of diesel and fuel oil in soil and aquatic habitats. Higher abundance (up to 30% of sequences) of *Corynebacterium,* including *Auritibacter, Brevibacterium* and *Cutibacterium* have been recorded in saline activated sludge, influent and effluent wastewater (Ye and Zhang, 2011). Members of the family

Corynebacteriaceae are potentially pathogenic bacterial groups generally associated with humans and animals (Ye and Zhang, 2011; Zhang et al., 2019).

On the other hand, *Proteobacteria,* a large bacterial phyla with versatile metabolic capabilities, are significantly enriched in river sediments exposed to inflow of wastewater treatment plants (WWTPs) effluents rich in organic compounds (Godoy et al., 2020; McLellan et al., 2015; Nega et al., 2019; Zhang et al., 2019) and heavy metals pollution (Luo and Ma, 2018; Ni et al., 2016; Zhao et al., 2020). These studies reported that these bacteria are crucial for the breakdown of organic matter and the cycling of nutrients. Congruent to current findings, members of the metal-tolerant *Gammaproteobacteria* have previously been reported to show high abundance in ecosystem polluted with heavy metals such as Zn, Pb and Cd (Zhao *et al*., 2020). The significant enrichment of phylum *Actinobacteria* and *Proteobacteria*, indicates their collective important role in the elimination of various contaminants, such as carbon, nitrogen, and phosphorous and heavy metals in the polluted river sediments (mid and lower catchment zones). Members of phylum *Verrucomicrobia* are ubiquitous in nature exhibiting a cosmopolitan distribution in freshwater lakes and rivers, and are suggested to play a key role as polysaccharide degraders (He *et al*., 2017). However, they have been reported to be very sensitive to heavy metal concentrations and cannot adapt to heavy metal polluted environments (Zhao *et al*., 2020). Herein, *Verrucomicrobia* was significantly enriched in the non-polluted upper catchment than polluted mid and lower river catchment. Therefore, their negative relationships with heavy metals concentrations can explain their lower abundances in both lower and mid river catchment sediments.

5.1.2 Core microbiome and taxonomic biomarkers of stream sediments

Overall, there were 3 core OTUs (unclassified *Corynebacteriaceae*, *Rhizomicrobium* and *Staphylococcus*) accounting for 18.9% of the total OTUs that were shared among the three catchments (Figure 4.4b). Members of family *Corynebacteriaceae* and genus *Staphylococcus* are human and animal-associated potentially pathogenic bacterial groups (Ye and Zhang, 2011; Zhang et al., 2019), while *Rhizobium* are important nitrifiers in freshwater ecosystems (Fisher *et al*., 2015) and urban river affected by agricultural activities and wastewater pollution (Godoy et al., 2020; Huang et al., 2018; Zhang et al., 2019). Differences in variations in the unique OTUs were observed between the catchment zones. Curiously, mid and lower catchments had 4 (1.2%) and 19 (5.2%) unique OTUs, respectively; the bulk of these taxa being human- and infrastructure-

associated bacteria related to greywater and wastewater contamination (Benami et al., 2013; Mishra and Mohanraju, 2018). These results, therefore, indicates the highly urbanized lower and mid catchment of the streams flowing through Kisumu City suffers from anthropogenic pollution mainly associated with wastewater discharges and greywater contamination.

Upper and mid catchment zones had the largest number of taxonomic biomarkers. Genus *Cellulomonas, Lysobacter, Ricketsia, unclassified Xanthobacteriaeceae and unclassified Bacillaceae* were the key taxonomic biomarkers in the upper zone (Figure 4.5a). Similarly, five taxa (*Corynebacterium, Cutibacterium, Staphylococcus, Bravibacterium* and *unclassified Micrococcaceae*) that have been linked to grey water pollution in urban ecosystems (Benami et al., 2013; Mishra and Mohanraju, 2018), were detected as the key taxonomic biomarkers for the mid catchment zone. In contrast, only genus *Ensifer* and *Amycolatopsis* were the main taxonomic biomarkers in lower catchment zone. Consistent with the findings of this study, *Ensifer* (*Sinorhizobium*) are known efficient heavy metals oxidizing (diCenzo *et al*., 2018) and a common nitrogen fixing bacteria in wastewater sludge (Ben Rebah et al., 2002). Similarly, *Amycolaptopsis* is heavy-metals (Pb, Cr, Zn and Cu) tolerant and nitrile degrading bacterium related to urban pollution (Ni *et al*., 2016; Zhao *et al*., 2020).

5.2 Functional profile of sediment bacterial communities and potential pathogens distribution

The three stream catchment sediments had comparable abundance of pathways related to metabolism (carbon, sulfur, nitrogen and methanol) and disease pathogenesis. However, significant differences ($p < 0.05$) in the abundance of key pathways were identified, mainly related to nitrogen metabolism and pathogenesis. Whereas, disease/virulence markers such as polymyxin resistance, peptidoglycan biosynthesis V (beta-lactam resistance) and II (staphylococci), and nitrifier denitrification pathway were significantly enriched in lower and mid catchment zones, upper zone sediment was characterized by higher assimilatory nitrate reduction I and urea cycle pathways (Figure 4.6a). Members of pathogenic staphylococci (e.g. *Stapylococcus aureus*) are renowned for their resistance to many commonly used antibiotics and prevalence in hospitals attributed to virulence marker peptidoglycan biosynthesis II (Reed et al., 2015). *Mycobacteria* and *Enterococcus,* whose infectivity can be related to peptidoglycan biosynthesis III and IV pathways, respectively, faecal-oral transmission has been reported to occur through sewage contamination of freshwater sources (Mtetwa et al., 2022; Young et al., 2016). In contrast, antibiotic resistance

pathway such as polymyxin resistance has been associated with virulence and pathogenesis of gram-negative pathogens such as *E. coli*, *Klebsiella pneumoniae*, and *K. oxytoca* isolated from selected sewage polluted urban rivers in Addis Ababa, Ethiopia (Belachew et al., 2018).

There have been numerous reports of significant pathogen contamination of urban surface waterways in developing nations, which has resulted in outbreaks of a number of waterborne diseases (Cai and Zhang, 2013; Godoy et al., 2020; Islam et al., 2020; Patel et al., 2016). However, high throughput sequencing has only seldom been used in studies of waterborne pathogens in urban surface waters, particularly in sub-Saharan Africa. Ten potential pathogenic bacterial genera, many of which are connected to human stool, animal feces (especially swine and cattle), grey water, and household WWTPs (Godoy et al., 2020; McLellan et al., 2015; Nega et al., 2019; Zhang et al., 2019), showed considerable differential enrichment. Overall, the putative pathogens distribution in the contaminated lower and mid catchment sediments showed increased alpha diversity, implying that the local potential pathogen composition was affected by the pollution inputs. The key putative pathogenic taxa enriched in the lower and mid catchment sediments included *Corynebacterium*, *Staphylococcus*, *Cutibacterium*, *Turicella*, *Acinetobacter*, *Micrococcus* and *Faecalibacterium*. Compared with other catchment, upper zone had significant enrichment ($p < 0.05$) of *Bacillus*, *Clostridium* and *Acidovorax*, whereas *Methylobacterium* and *Mycobacterium* exhibited significant enrichment ($p < 0.05$) in the lower zone (Figure 4.6b). Other important potential pathogenic bacteria detected in lower and mid catchment zones, albeit at low abundance, included the key gut-associated bacterial genera such as *Shewanella*, *Escherichia*, *Klebsiella*, *Enterococcus*, *Prevotella*, *Legionella*, *Vibrio*, *Salmonella* and urban wastewater-rich genera such as *Campylobacter, Roseomonas, Burkholderia* and *Acetobacterium.* Interestingly, enhanced virulence and pathogenesis biomarkers have previously been connected with potential pathogen dispersal, suggesting that urbanization poses a severe emerging risk for the human infections dissemination, propagation, and transmission as well as antibiotic resistance (Numberger *et al.*, 2021).

The findings of this study are important as the urban population in Kisumu City is continuously exposed to these uniquely hazardous environments that could increase their risk of infection by multiple pathogen types. Moreover, river water subject to wastewater contamination is often used for washing of clothes and food utensils and for bathing and even cooking. Results from the current study corroborates previous work that have reported that neighborhood landscapes in Kisumu, Kenya, especially urban streams are heavily contaminated by many enteric

viral, bacterial, protozoan, and helminthic enteric pathogen species (Baker et al., 2018), with records from health facilities showing that diarrheal diseases are among the top causes of morbidity. In addition, high levels of environmental contamination, often associated with improper waste and excreta management, are widespread among informal settlements within urban areas that contaminates public water sources (Opisa et al., 2012; Rochelle-Newall et al., 2015). Thus, adequate structures are urgently needed for the long-term monitoring of water borne pathogens in the environment in Kisumu City and other rapidly growing cities in sub–Saharan Africa. Furthermore, increasing the knowledge base on the dynamics of the water borne pathogens in river ecosystems is needed to be able to reduce the risks associated with the use of untreated water in Kisumu City. The results of this study suggest that sediment bacterial communities may be used widely across different catchment zones, irrespective of sample location, to monitor the effects of pollution on the ecological health of urban rivers and streams across sub-Saharan Africa. Interestingly, sediment samples from polluted river sections clustered distinctively and displayed a higher bacterial diversity than non-polluted zones, strengthening the fact that sediments may serve as reservoirs of diverse bacterial populations. This highlights the need to include sediments microbiome analysis for continuous monitoring of river/stream health.

5.3 Physicochemical characteristics and heavy metal pollution of stream sediments

In this study, results showed that the lower and mid catchment zone sediments had significant ($p < 0.01$) elevated levels of TOC, TN, and TP than upper catchment zone (Table 4.2). In addition, lower and mid catchment zone sediment samples had significantly lower ($p < 0.05$) pH values than upper catchment zone. The peri-urban upper catchment sediments samples exhibited two- to ten-fold lower heavy metal contents than other sampling sites, indicating that river water pollution is mainly linked to urban anthropogenic inputs. Consistent with this study, Outa et al (2020) reported higher levels of TOC, TN and TP and heavy metals (Zn, Pb, As, Cu, Cd and Cr) in the surface water and sediments at the inlets of the Kisat river into Lake Victoria. The urban segments of the Msimbazi River in Dar es Salaam, Mirongo River in Mwanza, Imeta River in Mbeya, and Ngerengere-Morogoro River in Morogoro, Tanzania, have previously observed declines in river water quality caused by enrichment of TP, PO_4^{3-}, NH_4^+, COD_{Mn}, and NO_3 (Chen et al., 2022). Heavy metal concentrations exceeding the values ($mg.kg^{-1}$) of 180.3 (Cu), 451.5 (Zn), 185.8 (Pb), and 4.1 (Hg) have been in sediment samples from the N'djili River, which drains through Kinshasa City, Democratic Republic of the Congo, according to Tshibanda *et al*

(2021). According to these investigations, both direct and diffuse sources of contamination were responsible for the pollution of urban rivers.

Domestic raw sewage and grey wastewater runoff from low-income cities in Africa (Chen et al., 2022; Nyilitya et al., 2016; Outa et al., 2020; Van der Hoven et al., 2017) and elsewhere (Li et al., 2022; Moncayo-Estrada et al., 2017; Tromboni and Dodds, 2017) have been linked to anthropogenic C and NH_4^+-N pollution with increasing population growth. In addition, detergent use and household food consumption make up the primary sources for TP pollution from urban households (Xiong *et al*., 2020). While anthropogenic C and NH_4^+-N pollution with increasing population growth have been linked to direct discharge of domestic raw sewage and grey wastewater or via runoff in low-income cities in Africa and elsewhere, detergent use and household food consumption constitute primary sources for TP pollution from urban households in SSA (Xiong *et al*., 2020). In this study, the lower catchment zone included the heavily built central business district characterized by presence of several industries and well-built residential estates in the outskirts, serviced with two wastewater treatment plants and a large open solid waste dumpsite. The zone is also characterized by heavy vehicular activities and high concentration of open vehicle repairs garages and carwashes. Abrasion of vehicle tyres, brake discs and other parts of vehicles are common sources of Cd pollution of soils and water (Adamiec et al., 2016). Therefore, possible sources of Pb and Cd pollution in the rivers and Lake Victoria include runoff from polluted roadside soils and open garages and carwashes (with Pb-based paint scrapings and Cd-rich vehicle part emissions) (Outa *et al*. 2020). Additionally, the discharge of treated and untreated wastewater from commercial and industrial communities, as well as densely populated places, may be a cause in the organic pollution found in the river sediments from the low catchment. In contrast to lower catchment zone, the mid catchment zone is mainly residential, with proliferation of informal settlements (slum) that are generally not serviced with piped sewerage system and formal solid waste management. The key diffuse sources of river pollution in this zone included untreated sewage, grey water, surface runoff, uncontrolled landfills and carwashes in riverbanks, that greatly contributes to both organic and inorganic pollutants loading. For example, greywater is also a rich source of organic and inorganic pollutants which include heavy metals such as Mn, Cu, Zn, Fe, Pb, Cr, Ni, Co and Cd (Chiroma et al., 2014; Katukiza et al., 2015; Li et al., 2021) and potential pathogens (Benami *et al*., 2013).

Based on the Nemerow pollution assessment index (P_N), highly urbanized catchment zones exhibited severe heavy metal pollution, mainly attributed to Pb, Cd and Cu severe pollution in the

lower catchment, and Pb severe pollution in the mid catchment (Figure 4.7). The heavy metal pollution level was comparable to those reported in urban rivers (Allafta and Opp, 2020; Tshibanda et al., 2021), but much lower than urban water ecosystem impacted by intensive mining and industrial pollution (Dey *et al*., 2021; Godoy *et al*., 2020). With increasing urbanization, therefore, heavy metal pollution of rivers in Kisumu City is expected to worsen, and thus need to be addressed urgently for improved environmental and public health.

5.4 Environmental drivers of sediment microbiome variation

There is evidence that the dynamics of bacterial populations in surface water and sediment ecosystems are influenced by a number of physicochemical factors, including TOC, TP, NO_3^-, and heavy metals (Zn, Fe, and Pb) concentrations (Fisher et al., 2015; Medeiros et al., 2016; Wang et al., 2018; Zhang et al., 2019). The main factors that grouped the lower and mid catchment samples were TOC, Pb, Cd, TN and Cr, key indicators of urbanized anthropogenic pollution. Meanwhile, cluster of upper catchment samples appeared to be mostly influenced by pH (Figure 4.8). The role of pH as an integrating variable that offers an integrated measure of the soil and water environment that influences bacterial populations has been noted in a number of research (Liu *et al*., 2015). One general explanation for why river sediment pH was the best predictor of community composition in the non-polluted upper catchment is that, soil or sediment pH may directly or indirectly affect factors such as nutrient availability, cationic metal solubility, organic C characteristics, soil moisture regime, and salinity (Lauber et al., 2009; Liu et al., 2015; Shen et al., 2013); factors that may be responsible for the observed changes in the community composition. In contrast, MonteCarlo permutation test showed that the factor with the most significant influence on clustering of the bacterial populations to the lower and mid zones was Pb ($p < 0.01$), followed by TOC ($p < 0.01$) and Cd content ($p < 0.01$). In support of this findings, linear regression analysis also revealed significant positive moderate correlation ($R > 0.4$, $p < 0.01$) between Pb and TOC content with microbiome Bray-Curtis distances (Figure 4.8b).

The amount and the composition of TOC have previously been reported as key determinant of archaeal and bacterial community structure in the surface and deeper sediments of Helgoland mud area in North Sea (Oni et al., 2015), where they may serve as microbial energy sources in methanic subsurface environments. Consistent to current findings, long-term agricultural and sewage pollution of rivers have also been reported to increase significantly the TOC, TP and TN

content in river sediments, that affects the distribution of microbial communities in river sediments (Zhang et al., 2021), with TP and TN levels affecting the proportion of *Proteobacteria*. In contrast to the current findings, both Pb and Cd have been reported to show a uniformly negative associations with the relative abundance of *Nitrospirae, Bacteroidetes* and *Verrucomicrobia*; including affecting predicted functions of microbial communities, such as metabolic functions, genetic information processes, and functions related to the carbon cycle and the nitrogen cycle. Overall, the results of the microbial community structure (Figure 4.2) and the analysis of microbial variations at the genus level (Figure 4.4) and RDA results (Figure 4.8), HMs and organic carbon inputs associated with urban pollution are the key drivers of river sediment bacterial community composition and structure. Therefore, there is an urgent need to improve the urban infrastructure and sanitation management related to both sewage treatment and solid waste management to curb pollution related to urban rivers/streams.

5.5 Limitations of the study

Despite the study providing valuable information for ecological risk assessment and management of urban rivers exposed to different types of pollutants, in addition to illustrating the potential use of sediment bacterial communities in monitoring and controlling urban pollution of surface waters, the current study had several limitations:

1. The time frame of samples collection was too short to observe the change in bacterial communities in urban impacted Kisat and Auji streams under long term exposure to the contaminants. Therefore, a more comprehensive and long-term study (covering more than two seasons and years) of sediment bacterial communities will be important to provide full insight into the monitoring and control of contamination in urban rivers.

2. This study's findings only provide insight into the taxonomy and phylogeny of bacterial communities as influenced by both organic and inorganic pollutants, but it did not identify the attendant genes that allow the bacterial communities to survive in such an ecosystem.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The key findings of this study included:

1. This study provided an in-depth insight of the link between urban pollution and sediment microbial community in sediment and potential bacterial pathogens. Despite exhibiting similarly taxonomic profile consisting of *Proteobacteria, Actinobacteria, Chloroflexi* and *Verrucomicrobia,* sediment microbial abundances significantly varied between upper, mid and lower catchments. Linear Discriminant Analysis Effect Size (LEfSe) detected five taxa (*Corynebacterium, Cutibacterium, Staphylococcus, Bravibacterium* and *unclassified Micrococcaceae*) associated with grey water pollution in urban ecosystems as the key taxonomic biomarkers for the mid catchment zone, whereas *Ensifer* (*Sinorhizobium*) and *Amycolatopsis,* known efficient heavy metal-tolerant oxidizers, nitrogen fixing bacteria and xenobiotic degrading bacteria, were the main taxonomic biomarkers in lower catchment zone.

2. Disease/virulence markers such as polymyxin resistance, peptidoglycan biosynthesis V and II and nitrifier denitrification pathway were significantly enriched in lower and mid catchment zones. Upper zone sediment was characterised by higher assimilatory nitrate reduction I and urea cycle pathways. *Mycobacteria* and *Enterococcus* infectivity were related to peptidoglycan biosynthesis III and IV pathways, respectively. The highly enriched human pathogenic genera identified included *Corynebacterium*, *Staphylococcus*, *Cutibacterium*, *Turicella*, *Acinetobacter* and *Faecalibacterium*, and gut-associated bacterial genera such as *Shewanella*, *Escherichia*, *Klebsiella*, *Enterococcus*, *Prevotella*, *Legionella*, *Vibrio*, and S*almonella* was detected in polluted mid and lower catchment zones. In addition, other human-associated potential bacteria pathogens such as *Bacillus*, *Streptococcus, Clostridium, Micrococcus*, *Haemophilus*, *Acidovorax*, *Veilonella*, *Chryseobacterium, Methylobacterium, Dermabacter*, *Abiotrophia* and *Mycobacterium* were also identified. These findings imply that human bacterial pathogens are dispersed via wastewater from domestic effluents and several anthropogenic activities that take place near the rivers, and thus constitute a serious public health in the study area.

Moreover, the persistent enrichment of candidate pathogens and faecal bacteria in sediment in polluted river catchments suggests that they may one day be significant biomarkers of the health and status of urban river ecosystems under pollution stress.

3. Water and sediments of Kisat, and Auji river exhibited variable organic and heavy metals pollution attributed to varying urbanization within their catchment. The average P_N of sediments for lower, mid and upper catchment zones were 4.97±0.98, 4.59±0.47 and 0.84±0.31, respectively, indicating that the lower and mid catchment zones were "severely polluted" ($P_N > 3.0$) with heavy metals. Specifically, the sediment and water samples mid and lower catchment zones of the two rivers exhibited higher organic (TOC, TN) pollution and severe Pb, Cd, and Cr pollution ($P_i > 3.0$) related to municipal wastewater effluent discharge, greywater contamination, industrial wastewater disposal and other point sources pollution from run-off of contaminated roadside soils and open garages and carwashes. Multivariate (RDA) analysis revealed that sediment bacterial community diversity and structure in the two rivers was significantly influenced by TOC, Pb, Cd, TN, Cr and pH with *Corynebacteriales, Micrococcales, Pseudonocardiales, Propionibacteriales* and *Bacillales* were the main bacterial order detected. Higher bacterial diversity and distinct clustering of sediment samples from polluted river sections from non-polluted zones, provide evidence that sediments may function as reservoirs of various microbial groups whose perturbations may be utilised in urban river pollution monitoring.

6.2 Recommendations and suggestions for further study

1. Although previous studies have investigated the effect of pollutants on physiochemical parameters of aquatic ecosystems, an ecological understanding of how organic and inorganic contaminants mainly the heavy metals and bacterial communities interact due to long term exposure is needed in future studies. To this end, it will be crucial to conduct a more thorough and prolonged investigation of sediment bacterial populations (encompassing more than two seasons and years) to provide insight on the monitoring and management of contamination in urban rivers at both the local and national scales. This will be important in understanding the dynamic patterns in bacterial communities for proper management of urban rivers and large water bodies *e.g*., Lake Victoria.

2. This study showed the distribution of bacterial communities among the three catchment rivers as influenced by pollutants. Pollutants lead to change in physiochemical parameters of aquatic ecosystem which eventually change their bacterial composition majority of which are potential human pathogens. However, previous studies have shown pollutants can induce adaptive variation in the genetic make-up of bacterial communities which may lead to mutations (Najafi & Pezeshki, 2013). According to Gao *et al*., 2021, bacterial communities are able to develop resistance to heavy metal pollution. It is therefore, important for future studies to unravel the attendant genes that enable these biological communities to survive in polluted environments.

3. Future studies should incorporate other physicochemical parameters not included in this study but might be important environmental factors influencing sediment bacterial community composition and structure in urban-impacted rivers/streams.

REFERENCES

Adamiec, E., Jarosz-Krzemińska, E., Wieszała, R., 2016. Heavy metals from non-exhaust vehicle emissions in urban and motorway road dusts. Environ. Monit. Assess. 188, 369. https://doi.org/10.1007/s10661-016-5377-1

Ahmed, W., Staley, C., Sadowsky, M.J., Gyawali, P., Sidhu, J.P.S., Palmer, A., Beale, D.J., Toze, S., 2015. Toolbox Approaches Using Molecular Markers and 16S rRNA Gene Amplicon Data Sets for Identification of Fecal Pollution in Surface 81, 7067–7077. https://doi.org/10.1128/AEM.02032-15

Allafta, H., Opp, C., 2020. Spatio-temporal variability and pollution sources identification of the surface sediments of Shatt Al-Arab River, Southern Iraq. Sci. Rep. 10, 6979. https://doi.org/10.1038/s41598-020-63893-w

Ancion, P.-Y., Lear, G., Lewis, G.D., 2010. Three common metal contaminants of urban runoff (Zn, Cu & Pb) accumulate in freshwater biofilm and modify embedded bacterial communities. Environ. Pollut. 158, 2738–2745. https://doi.org/https://doi.org/10.1016/j.envpol.2010.04.013

Andersen, K.S., Kirkegaard, R.H., Albertsen, M., 2018. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv. https://doi.org/http://dx.plos.org/10.1371/journal.pone.0132783

Anderson, M.J., Walsh, D.C.I., 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? Ecol. Monogr. 83, 557–574.

Andreasen, M.H., Agergaard, J., Kiunsi, R.B., Namangaya, A.H., 2017. Urban transformations, migration and residential mobility patterns in African secondary cities. Geogr. Tidsskr. J. Geogr. 117, 93–104.

Aneyo, I.A., Funmilayo, V., Adebesin, O.A., Hammed, M.O., 2016. Biodegradation of Pollutants in Waste Water from Pharmaceutical , Textile and Local Dye Effluent in Lagos , Nigeria. J. Heal. Pollut. 6, 34–42.

Angel, S., Blei, A.M., Civco, D.L., Parent, J., 2012. Atlas of urban expansion. Lincoln Institute

of Land Policy Cambridge, MA.

Bacosa, H.P., Inoue, C., 2020. Heavy oil degrading Burkholderia and Pseudomonas strains : insights on the degradation potential of isolates and microbial consortia. Palawan Sci. 12, 74–89.

Baker, K.K., Senesac, R., Sewell, D., Sen Gupta, A., Cumming, O., Mumma, J., 2018. Fecal fingerprints of enteric pathogen contamination in public environments of Kisumu, Kenya, associated with human sanitation conditions and domestic animals. Environ. Sci. Technol. 52, 10263–10274. https://doi.org/10.1021/acs.est.8b01528

Beattie, R.E., Bandla, A., Swarup, S., Hristova, K.R., 2020. Freshwater Sediment Microbial Communities Are Not Resilient to Disturbance From Agricultural Land Runoff. Front. Microbiol. 1, 539921. https://doi.org/10.3389/fmicb.2020.539921

Belachew, T., Mihret, A., Legesse, T., Million, Y., Desta, K., 2018. High level of drug resistance by gram-negative bacteria from selected sewage polluted urban rivers in Addis Ababa, Ethiopia. BMC Res. Notes 11, 524. https://doi.org/10.1186/s13104-018-3622-0

Ben Rebah, F., Tyagi, R.D., Prévost, D., 2002. Wastewater sludge as a substrate for growth and carrier for rhizobia: the effect of storage conditions on survival of Sinorhizobium meliloti. Bioresour. Technol. 83, 145–151. https://doi.org/https://doi.org/10.1016/S0960-8524(01)00202-4

Benami, M., Gross, A., Herzberg, M., Orlofsky, E., Vonshak, A., Gillor, O., 2013. Assessment of pathogenic bacteria in treated graywater and irrigated soils. Sci. Total Environ. 458, 298–302. https://doi.org/https://doi.org/10.1016/j.scitotenv.2013.04.023

Birks, R., Hills, S., 2007. Characterisation of Indicator Organisms and Pathogens in Domestic Greywater for Recycling. Environ. Monit. Assess. 129, 61–69. https://doi.org/1https://doi.org/10.1007/s10661-006-9427-y

Bodnar, I., Szabolcsik, A., Baranyai, E., Uveges, A., Boros, N., 2014. Qualitative characterization of household greywater in the northern great plain region of Hungary. Environ. Eng. Manag. J. 13, 2717–2724.

Bomar, L., Maltz, M., Colston, S., Graf, J., 2011. Directed Culturing of Microorganisms Using Metatranscriptomics. MBio 2, e00012-11. https://doi.org/https://doi.org/10.1128/mBio.00012-11

Breton-deval, L., Sanchez-reyes, A., Sanchez-flores, A., Ju, K., 2020. Functional Analysis of a Polluted River Microbiome Reveals a Metabolic Potential for Bioremediation. Microorganisms 8, 554. https://doi.org/https://doi.org/10.3390/microorganisms8040554

Brito, E.M.S., Piñón-Castillo, H.A., Guyoneaud, R., Caretta, C.A., Gutiérrez-Corona, J.F., Duran, R., Reyna-López, G.E., Nevárez-Moorillón, G.V., Fahy, A., Goñi-Urriza, M., 2013. Bacterial biodiversity from anthropogenic extreme environments: A hyper-alkaline and hyper-saline industrial residue contaminated by chromium and iron. Appl. Microbiol. Biotechnol. 97, 369–378. https://doi.org/10.1007/s00253-012-3923-5

Bryanskaya, A. V, Malup, T.K., Lazareva, E. V, Taran, O.P., Rozanov, A.S., Efimov, V.M., Peltek, S.E., 2016. The role of environmental factors for the composition of microbial communities of saline lakes in the Novosibirsk region 16, 1–15. https://doi.org/10.1186/s12866-015-0618-y

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. https://doi.org/10.1021/es400275r

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Jeremy, E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., 2011. NIH Public Access 7, 335–336. https://doi.org/10.1038/nmeth.f.303.QIIME

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 1621–1624. https://doi.org/10.1038/ismej.2012.8

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S

ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods 69, 330–339. https://doi.org/10.1016/j.mimet.2007.02.005

Chaudhary, A., Kauser, I., Ray, A., Poretsky, R., 2018. Taxon-Driven Functional Shifts Associated with Storm Flow in an Urban Stream Microbial Community. mSphere 3, e00194-18. https://doi.org/10.1128/mSphere.00194-18

Chen, S.S., Kimirei, I.A., Yu, C., Shen, Q., Gao, Q., 2022. Assessment of urban river water pollution with urbanization in East Africa. Environ. Sci. Pollut. Res. 29, 40812–40825. https://doi.org/10.1007/s11356-021-18082-1

Chiroma, T.M., Ebewele, R.O., Hymore, F.K., 2014. Comparative Assessement Of Heavy Metal Levels In Soil , Vegetables And Urban Grey Waste Water Used For Irrigation In Yola And Kano. Int. Ref. J. Eng. Sci. 3, 1–9.

Cui, Q., Huang, Y., Wang, H., Fang, T., 2019. Diversity and abundance of bacterial pathogens in urban rivers impacted by domestic sewage. Environ. Pollut. 249, 24–35. https://doi.org/https://doi.org/10.1016/j.envpol.2019.02.094

Davidson, C.M., 2013. Methods for the Determination of Heavy Metals and Metalloids in Soils BT - Heavy Metals in Soils: Trace Metals and Metalloids in Soils and their Bioavailability, in: Alloway, B.J. (Ed.), . Springer Netherlands, Dordrecht, pp. 97–140. https://doi.org/10.1007/978-94-007-4470-7_4

Dey, M., Akter, A., Islam, S., Chandra Dey, S., Choudhury, T.R., Fatema, K.J., Begum, B.A., 2021. Assessment of contamination level, pollution risk and source apportionment of heavy metals in the Halda River water, Bangladesh. Heliyon 7, e08625. https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e08625

diCenzo, G.C., Debiec, K., Krzysztoforski, J., Uhrynowski, W., Mengoni, A., Fagorzi, C., Gorecki, A., Dziewit, L., Bajda, T., Rzepa, G., Drewniak, L., 2018. Genomic and Biotechnological Characterization of the Heavy-Metal Resistant, Arsenic-Oxidizing Bacterium Ensifer sp. M14. Genes (Basel). 9, 379. https://doi.org/10.3390/genes9080379

Domingo, J.W.S., Edge, T.A., 2010. Identification of primary sources of faecal pollution.

Douglas, G.M., Beiko, R.G., Langille, M.G.I., 2018. Predicting the Functional Potential of the Microbiome from Marker Genes Using PICRUSt BT - Microbiome Analysis: Methods and Protocols, in: Beiko, R.G., Hsiao, W., Parkinson, J. (Eds.), . Springer New York, New York, NY, pp. 169–177. https://doi.org/10.1007/978-1-4939-8728-3_11

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

Enander, F., 2017. Elemental distribution in the catchment around Lake Victoria in Kenya. Linköping University.

Fashola, M.O., Ngole-jeme, V.M., Babalola, O.O., 2016. Heavy Metal Pollution from Gold Mines : Environmental Effects and Bacterial Strategies for Resistance. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph13111047

Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elem. (Washington, D.C.) 3, 64. https://doi.org/10.12952/journal.elementa.000064

Fouad, A.F., 2012. Pyrosequencing as a tool for better understanding of human microbiomes 2297. https://doi.org/10.3402/jom.v4i0.10743

Godoy, R.G., Marcondes, M.A., Pessôa, R., Nascimento, A., Victor, J.R., Duarte, A.J. da S., Clissa, P.B., Sanabani, S.S., 2020. Bacterial community composition and potential pathogens along the Pinheiros River in the southeast of Brazil. Sci. Rep. 10, 1–9. https://doi.org/10.1038/s41598-020-66386-y

Gross, A., 2007. Recycled vertical flow constructed wetland (RVFCW)— a novel method of recycling greywater for irrigation in small communities and households. Chemosphere 66, 916–923. https://doi.org/10.1016/j.chemosphere.2006.06.006

Han, J.L., Pan, X.D., Chen, Q., Huang, B.F., 2021. Health risk assessment of heavy metals in marine fish to the population in Zhejiang ,. Sci. Rep. 1–9. https://doi.org/10.1038/s41598-021-90665-x

Haruta, S., Nanako, K., 2015. Survivability of Microbes in Natural Environments and Their

Ecological Impacts. Microbes Environ. 30, 123–125.
https://doi.org/10.1264/jsme2.ME3002rh

Harwood, V.J., Staley, C., Badgley, B.D., Borges, K., Korajkic, A., 2013. Microbial source
tracking markers for detection of fecal contamination in environmental waters: relationships
between pathogens and human health outcomes 1–40. https://doi.org/10.1111/1574-
6976.12031

He, S., Stevens, S.L.R., Chan, L.-K., Bertilsson, S., del Rio, T.G., Tringe, S.G., Malmstrom,
R.R., McMahon, K.D., 2017. Ecophysiology of freshwater Verrucomicrobia inferred from
metagenome-assembled genomes. bioRxiv 150078. https://doi.org/10.1101/150078

Hirai, J., Nagai, S., Hidaka, K., 2017. Evaluation of metagenetic community analysis of
planktonic copepods using Illumina MiSeq : Comparisons with morphological classification
and metagenetic analysis using Roche 454 1–19.

Hosen, J.D., Febria, C.M., Crump, B.C., Palmer, M.A., 2017. Watershed Urbanization Linked to
Differences in Stream Bacterial Community Composition. Front. Microbiol.
https://doi.org/10.3389/fmicb.2017.01452

Hu, B., Ruiying, Z., Chen, S., Zhou, Y., Jin, B., Li, Y., Shi, Z., 2018. Heavy Metal Pollution
Delineation Based on Uncertainty in a Coastal Industrial City in the Yangtze River Delta,
China. Int. J. Environ. Res. Public Health 15, 710. https://doi.org/10.3390/ijerph15040710

Huang, S., Chen, C., Jaffé, P.R., 2018. Seasonal distribution of nitrifiers and denitrifiers in urban
river sediments affected by agricultural activities. Sci. Total Environ. 642, 1282–1291.
https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.06.116

Iloms, E., Ololade, O.O., Ogola, H.J.O., Selvarajan, R., 2020. Investigating Industrial Effluent
Impact on Municipal Wastewater Treatment Plant in Vaal, South Africa. Int. J. Environ.
Res. Public Health 17, 1096. https://doi.org/10.3390/ijerph17031096

Islam, Md. Sirajul, Hassan-uz-Zaman, M., Islam, Md. Shafiqul, Clemens, J.D., Ahmed, N., 2020.
Chapter 3 - Waterborne pathogens: review of outbreaks in developing nations, in: Vara
Prasad, M.N., Grobelak, A.B.T.-W.P. (Eds.), . Butterworth-Heinemann, pp. 43–56.
https://doi.org/https://doi.org/10.1016/B978-0-12-818783-8.00003-7

Jacquiod, S., Cyriaque, V., Riber, L., Al-soud, W.A., Gillan, D.C., Wattiez, R., Sørensen, S.J., 2018. Long-term industrial metal contamination unexpectedly shaped diversity and activity response of sediment microbiome. J. Hazard. Mater. 344, 299–307. https://doi.org/https://doi.org/10.1016/j.jhazmat.2017.09.046

James, D.T.K., Surendran, S., Ifelebuegu, A.O., Ganjian, E., Kinuthia, J., 2016. Grey water reclamation for urban non-potable reuse–challenges and solutions: a review, in: Proceeding 7th International Conference on Sustainable Built Environment at Sri Lanka (Kandy).

Jefferson, B., Laine, A., Parsons, S., Stephenson, T., Judd, S., 2000. Technologies for domestic wastewater recycling. Urban water 1, 285-292.

Jefferson, B., Palmer, A., Jeffrey, P., Stuetz, R., Judd, S., 2004. Grey water characterisation and its impact on the selection and operation of technologies for urban reuse. Water Sci. Technol. 50, 157–164.

Jiang, H., Dong, H., Zhang, G., Yu, B., Chapman, L.R., Fields, M.W., 2006. Microbial Diversity in Water and Sediment of Lake Chaka , an Athalassohaline Lake in Northwestern China. Appl. Environ. Microbiol. 72, 3832–3845. https://doi.org/10.1128/AEM.02869-05

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. 10, 5029. https://doi.org/10.1038/s41467-019-13036-1

Jordaan, K., Bezuidenhout, C.C., 2013. The impact of physico-chemical water quality parameters on bacterial diversity in the Vaal River, South Africa. Water Sa 39, 385–396.

Jordaan, K., Comeau, A.M., Khasa, D.P., Bezuidenhout, C.C., 2019. An integrated insight into the response of bacterial communities to anthropogenic contaminants in a river : A case study of the Wonderfonteinspruit catchment area , South Africa. PLoS One 14, e0216758. https://doi.org/10.1371/journal.pone.0216758

K'oreje, K.O., Vergeynst, L., Ombaka, D., De Wispelaere, P., Okoth, M., Van Langenhove, H., Demeestere, K., 2016. Occurrence patterns of pharmaceutical residues in wastewater, surface water and groundwater of Nairobi and Kisumu city, Kenya. Chemosphere 149, 238–

244. https://doi.org/https://doi.org/10.1016/j.chemosphere.2016.01.095

Kapembo, M.L., Al Salah, D.M.M., Thevenon, F., Laffite, A., Bokolo, M.K., Mulaji, C.K., Mpiana, P.T., Poté, J., 2019. Prevalence of water-related diseases and groundwater (drinking-water) contamination in the suburban municipality of Mont Ngafula, Kinshasa (Democratic Republic of the Congo). J. Environ. Sci. Heal. Part A 1–11.

Kassambara, A., 2020. ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.4.0.

Katukiza, A.Y., Ronteltap, M., Niwagaba, C.B., Kansiime, F., Lens, P.N.L., 2015. Grey water characterisation and pollutant loads in an urban slum 423–436. https://doi.org/10.1007/s13762-013-0451-5

Kayaga, S., Njiru, C., Itiko, M., Onyango-Awin, J., 2005. Improving utility management: case study from Kisumu, Kenya., in: Maximizing the Benefits from Water and Environmental Sanitation: 31st WEDC Conference, Kampala, Uganda. Water, Engineering and Development Centre (WEDC) Loughborough University of ….

Kimosop, S.J., Getenga, Z.M., Orata, F., Okello, V.A., Cheruiyot, J.K., 2016. Residue levels and discharge loads of antibiotics in wastewater treatment plants (WWTPs), hospital lagoons, and rivers within Lake Victoria Basin, Kenya. Environ. Monit. Assess. 188, 1–9.

Kinuthia, G.K., Ngure, V., Beti, D., Lugalia, R., Wangila, A., 2020. Levels of heavy metals in wastewater and soil samples from open drainage channels in Nairobi , Kenya : community health implication. Sci. Rep. 10.1, 1–13. https://doi.org/10.1038/s41598-020-65359-5

KNBS, 2019. 2019 Kenya population and housing census volume II: Distribution of population by administrative units. Kenya National Bureau of Statistics, Nairobi.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent microbial source tracking. Nat. Methods 8, 761–763. https://doi.org/10.1038/nmeth.1650

Köchling, T., Sanz, J.L., Galdino, L., Florencio, L., Kato, M.T., 2017. Impact of pollution on the microbial diversity of a tropical river in an urbanized region of northeastern Brazil. Int.

Microbiol. 20, 11–24. https://doi.org/10.2436/20.1501.01.281.

Kolding, J., Medard, M., Mkumbo, O.C., van Zwieten, P.A.M., 2013. Status, trends and management of the Lake Victoria Fisheries, Inland fisheries evolution and management – case studies from four continents, FAO Fisheries and Aquaculture Technical Paper. Rome, Italy.

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. https://doi.org/10.1128/AEM.01043-13

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., Beiko, R.G., Huttenhower, C., 2013. A n a ly s i s Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814–821. https://doi.org/10.1038/nbt.2676

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75, 5111–5120.

Lee, J., Chon, H.T., Jung, M.C., 2005. Toxic risk assessment and environmental contamination of heavy metals around abandoned metal mine sites in Korea, in: Key Engineering Materials Vol. 277. Trans Tech Publications Ltd, pp. 542–547.

Li, H., Liang, S., Liang, Y., Li, K., Qi, J., Yang, X., Feng, C., Cai, Y., Yang, Z., 2021. Multi-pollutant based grey water footprint of Chinese regions. Resour. Conserv. Recycl. 164, 105202.

Li, Y., Fang, L., Yuanzhu, W., Mi, W., Ji, L., Guixiang, Z., Yang, P., Chen, Z., Bi, Y., 2022. Anthropogenic activities accelerated the evolution of river trophic status. Ecol. Indic. 136, 108584. https://doi.org/https://doi.org/10.1016/j.ecolind.2022.108584

Lightbody, G., Haberland, V., Browne, F., Taggart, L., Zheng, H., Parkes, E., Blayney, J.K., 2019. Review of applications of high-throughput sequencing in personalized medicine:

barriers and facilitators of future progress in research and clinical application. Brief. Bioinform. 20, 1795–1811. https://doi.org/10.1093/bib/bby051

Liu, S., Ren, H., Shen, L., Lou, L., Tian, G., Zheng, P., Hu, B., 2015. pH levels drive bacterial community structure in sediments of the Qiantang River as determined by 454 pyrosequencing. Front. Microbiol. https://doi.org/10.3389/fmicb.2015.00285

Lu, X., Zhang, X., Wang, Z., Huang, K., Wang, Y., Liang, W., 2015. Bacterial Pathogens and Community Composition in Advanced Sewage Treatment Systems Revealed by Metagenomics Analysis Based on High-Throughput Sequencing. PLoS One 10.5, 1–16. https://doi.org/10.1371/journal.pone.0125549

Luo, Z., Ma, J., 2018. Effects of Pb Smelting on the Soil Bacterial Community near a Secondary Lead Plant. Int. J. Environ. Res. Public Health 15.5, 1030. https://doi.org/10.3390/ijerph15051030

LVEMP, 2015. REHABILITATION AND EXTENSION OF THE WATER AND SANITATION NETWORK IN KISUMU: Environmental Impact Assessment Project Report. Kisumu.

Marais, L., Cloete, J., 2017. The role of secondary cities in managing urbanisation in South Africa. Dev. South. Afr. 34, 182–195.

Marcial Gomes, N.C., Borges, L.R., Paranhos, R., Pinto, F.N., Mendonça-Hagler, L.C.S., Smalla, K., 2008. Exploring the diversity of bacterial communities in sediments of urban mangrove forests. FEMS Microbiol. Ecol. 66, 96–109.

McLellan, S.L., Fisher, J.C., Newton, R.J., 2015. The microbiome of urban waters. Int. Microbiol. 18, 141–149. https://doi.org/10.2436/20.1501.01.244

Medeiros, J.D., Cantão, M.E., Cesar, D.E., Nicolás, M.F., Diniz, C.G., Silva, V.L., Vasconcelos, A.T.R. de, Coelho, C.M., 2016. Comparative metagenome of a stream impacted by the urbanization phenomenon. Braz. J. Microbiol. 47, 835–845. https://doi.org/10.1016/j.bjm.2016.06.011

Mikhailov, I.S., Zakharova, Y.R., Bukin, Y.S., Galachyants, Y.P., Petrova, D.P., Sakirko, M. V,

Likhoshway, Y. V, 2019. Co-occurrence networks among bacteria and microbial eukaryotes of Lake Baikal during a spring phytoplankton bloom. Microb. Ecol. 77, 96–109.

Mireji, P.O., Keating, J., Hassanali, A., Mbogo, C.M., Nyambaka, H., Kahindi, S., Beier, J.C., 2008. Heavy metals in mosquito larval habitats in urban Kisumu and Malindi, Kenya, and their impact. Ecotoxicol. Environ. Saf. 70, 147–153.

Mishra, A.K., Mohanraju, R., 2018. Epiphytic Bacterial Communities in Seagrass Meadows of Oligotrophic Waters of Andaman Sea. Open Access Libr. J. 05, 1–12. https://doi.org/10.4236/oalib.1104388

Moncayo-Estrada, R., Escalera-Gallardo, C., Arroyo-Damián, M., Campos-Campos, O., Silva-García, J.T., 2017. Nitrogen nutrients in a subtropical river: temporal variation and analysis at different spatial scales. Mar. Freshw. Res. 68, 1935–1949.

Mou, X., Lu, X., Jacob, J., Sun, S., Heath, R., 2013. Metagenomic Identification of Bacterioplankton Taxa and Pathways Involved in Microcystin Degradation in Lake Erie 8. https://doi.org/10.1371/journal.pone.0061890

Mtetwa, H.N., Amoah, I.D., Kumari, S., Bux, F., Reddy, P., 2022. The source and fate of Mycobacterium tuberculosis complex in wastewater and possible routes of transmission. BMC Public Health 22, 145. https://doi.org/10.1186/s12889-022-12527-z

Mukherjee, A., Chettri, B., Langpoklakpam, J.S., Basak, P., Prasad, A., Mukherjee, A.K., Bhattacharyya, M., Singh, A.K., Chattopadhyay, D., 2017. Bioinformatic Approaches Including Predictive Metagenomic Profiling Reveal Characteristics of Bacterial Response to Petroleum Hydrocarbon Contamination in Diverse Environments. Sci. Rep. 7, 1108. https://doi.org/10.1038/s41598-017-01126-3

Mukherjee, N., Bartelli, D., Patra, C., Chauhan, B. V, Dowd, S.E., Banerjee, P., 2016. Microbial Diversity of Source and Point-of-Use Water in Rural Haiti – A Pyrosequencing- Based Metagenomic Survey 1–17. https://doi.org/10.1371/journal.pone.0167353

Munala, G., Moirongo, B.O., 2017. The need for an integrated solid waste management in Kisumu, Kenya.

Mutuku, C., Paul, O., Hamadi, B., 2014. Metal pollutants distribution within Lake Victoria basin and their influence on the native and transient microbial flora. J. Agric. Biol. Sci. 9, 127–133.

Nega, M., Braun, B., Künzel, S., Szewzyk, U., 2019. Evaluating the Impact of Wastewater E ffl uent on Microbial Communities in the Panke, an Urban River. Water 11.5, 1–17. https://doi.org/10.3390/w11050888

Nemerow, N.L., 1974. Scientific stream pollution analysis.

Nguyen, T.T.N., 2017. Analysing the effect of industrial and urban polluted zones on microbial diversity in the SaiGon -DongNai river system (Vietnam). Université Paris-Saclay (ComUE).

Ni, C., Horton, D.J., Rui, J., Henson, M.W., Jiang, Y., Huang, X., Learman, D.R., 2016. High concentrations of bioavailable heavy metals impact freshwater sediment microbial communities. Ann. Microbiol. 1003–1012. https://doi.org/10.1007/s13213-015-1189-8

Numberger, D., Zoccarato, L., Woodhouse, J., Ganzert, L., Sauer, S., Grossart, H.-P., Greenwood, A., 2021. Urbanization promotes specific bacteria in freshwater microbiomes including potential pathogens. bioRxiv 2020.06.26.173328. https://doi.org/10.1101/2020.06.26.173328

Nyakeya, K., Raburu, P.O., Masese, F.O., Tsuma, J., Nyamora, J.M., Kemunto, D.D., Ondiba, R.N., Jepkosgei, M., 2017. Sensitivity of the Native Chironomus Species in Monitoring of Riverine Ecosystems in the Catchments of Lake Victoria Drainage Basin , Kenya. African Environ. Rev. J. 2, 126–133.

Nyilitya, B., Mureithi, S., Boeckx, P., 2016. Tracking sources of excess nitrate discharge in Lake Victoria, Kenya for improved Nitrogen use efficiency in the catchment, in: International Nitrogen Initiative Conference,"Solutions to Improve Nitrogen Use Efficiency for the World", Melbourne, Australia.

O'Toole, J., Sinclair, M., Malawaraarachchi, M., Hamilton, A., Barker, S.F., Leder, K., 2012. Microbial quality assessment of household greywater. Water Res. 46, 4301–4313. https://doi.org/10.1016/j.watres.2012.05.001

Ogola, H.J.O., Selvarajan, R., Tekere, M., 2021. Local Geomorphological Gradients and Land Use Patterns Play Key Role on the Soil Bacterial Community Diversity and Dynamics in the Highly Endemic Indigenous Afrotemperate Coastal Scarp Forest Biome. Front. Microbiol. 12, 281. https://doi.org/https://doi.org/10.3389/fmicb.2021.592725

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2010. Vegan: community ecology package. R package version 1.17-4. URL http//CRAN. R-project. org/package= vegan.

Okungu, J.O., Njoka, S., Abuodha, J.O.Z., Hecky, R.E., 2005. An introduction to Lake Victoria catchment, water quality, physical limnology and ecosystem status (Kenyan sector).

Olson, M., Harrington, C.T., Lin, E.I., Olson, M.T., Eshleman, J.R., 2009. Fundamentals of Pyrosequencing. Arch. Pathol. Lab. Med. https://doi.org/10.5858/arpa.2012-0463-RA

Oni, O.E., Schmidt, F., Miyatake, T., Kasten, S., Witt, M., Hinrichs, K.-U., Friedrich, M.W., 2015. Microbial Communities and Organic Matter Composition in Surface and Subsurface Sediments of the Helgoland Mud Area, North Sea . Front. Microbiol. .

Onyari, J.M., Wandiga, S.O., 1989. Distribution of Cr, Pb, Cd, Zn, Fe and Mn in lake Victoria sediments, East Africa. Bull. Environ. Contam. Toxicol. 42, 807–813. https://doi.org/10.1007/BF01701619

Opisa, S., Odiere, M.R., Jura, W.G.Z.O., Karanja, D.M.S., Mwinzi, P.N.M., 2012. Faecal contamination of public water sources in informal settlements of Kisumu City, western Kenya. Water Sci. Technol. 66, 2674–2681.

Outa, J.O., Kowenje, C.O., Plessl, C., Jirsa, F., 2020. Distribution of arsenic, silver, cadmium, lead and other trace elements in water, sediment and macrophytes in the Kenyan part of Lake Victoria: spatial, temporal and bioindicative aspects. Environ. Sci. Pollut. Res. 27, 1485–1498. https://doi.org/10.1007/s11356-019-06525-9

Pantha, K., Acharya, K., Mohapatra, S., Khanal, S., Amatya, N., Ospina-Betancourth, C., Butte, G., Shrestha, S.D., Rajbhandari, P., Werner, D., 2021. Faecal pollution source tracking in the holy Bagmati River by portable 16S rRNA gene sequencing. npj Clean Water 4, 12. https://doi.org/10.1038/s41545-021-00099-1

Patel, C.B., Shanker, R., Gupta, V.K., Upadhyay, R.S., 2016. Q-PCR based culture-independent enumeration and detection of Enterobacter: an emerging environmental human pathogen in riverine systems and potable water. Front. Microbiol. 7, 172. https://doi.org/https://doi.org/10.3389/fmicb.2016.00172

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Glo, F.O., Ludwig, W., 2007. SILVA : a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196. https://doi.org/10.1093/nar/gkm864

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Glo, F.O., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596. https://doi.org/10.1093/nar/gks1219

Reed, P., Atilano, M.L., Alves, R., Hoiczyk, E., Sher, X., Reichmann, N.T., Pereira, P.M., Roemer, T., Filipe, S.R., Pereira-Leal, J.B., Ligoxygakis, P., Pinho, M.G., 2015. Staphylococcus aureus Survives with a Minimal Peptidoglycan Synthesis Machine but Sacrifices Virulence and Antibiotic Resistance. PLoS Pathog. 11, e1004891–e1004891. https://doi.org/10.1371/journal.ppat.1004891

Reuter, J.A., Spacek, D. V, Snyder, M.P., 2015. High-Throughput Sequencing Technologies. Mol. Cell 58, 586–597. https://doi.org/https://doi.org/10.1016/j.molcel.2015.05.004

Rochelle-Newall, E., Nguyen, T.M.H., Le, T.P.Q., Sengtaheuanghoung, O., Ribolzi, O., 2015. A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions. Front. Microbiol. https://doi.org/https://doi.org/10.3389/fmicb.2015.00308

Saraceno, M., Gómez Lugo, S., Ortiz, N., Gómez, B.M., Sabio Y García, C.A., Frankel, N., Graziano, M., 2021. Unraveling the ecological processes modulating the population structure of Escherichia coli in a highly polluted urban stream network. Sci. Rep. 11, 14679. https://doi.org/10.1038/s41598-021-94198-1

Schloss, P.D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G.

Thallinger, D. J. Van Horn, C. F. Weber, 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., Von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science (80-.). 313, 1072–1077.

Segata, N., Huttenhower, C., 2011. Toward an Efficient Method of Identifying Core Genes for Evolutionary and Functional Microbial Phylogenies. PLoS One 6, e24704. https://doi.org/10.1371/journal.pone.0024704

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic biomarker discovery and explanation. Genome Biol. 12, R60–R60. https://doi.org/10.1186/gb-2011-12-6-r60

Shen, C., Xiong, J., Zhang, H., Feng, Y., Lin, X., Li, X., Liang, W., Chu, H., 2013. Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. Soil Biol. Biochem. 57, 204–211. https://doi.org/https://doi.org/10.1016/j.soilbio.2012.07.013

Shenhav, L., Thompson, M., Joseph, T.A., Briscoe, L., Furman, O., Bogumil, D., Mizrahi, I., Pe, I., Halperin, E., 2021. FEAST: fast expectation-maximization for microbial source tracking. Nat. Methods 16, 627–632. https://doi.org/https://doi.org/10.1038/s41592-019-0431-x

Sibanda, L.K., Obange, N., Awuor, F.O., 2017. Challenges of Solid Waste Management in Kisumu, Kenya. Urban Forum 28, 387–402. https://doi.org/10.1007/s12132-017-9316-1

Simpson, S.L., Batley, G.E., Chariton, A.A., 2007. Revision of the ANZECC/ARMCANZ sediment quality guidelines.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored '" rare biosphere ."' Proc. Natl. Acad. Sci. 103, 12115–12120. https://doi.org/https://doi.org/10.1073/pnas.0605127103

Sun, S., Jones, R.B., Fodor, A.A., 2019. Inference based PICRUSt accuracy varies across sample types and functional categories. bioRxiv 655746. https://doi.org/https://doi.org/10.1101/655746

Tripathi, B.M., Stegen, J.C., Kim, M., Dong, K., Adams, J.M., Lee, Y.K., 2018. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. ISME J. 12, 1072–1083. https://doi.org/10.1038/s41396-018-0082-4

Tromboni, F., Dodds, W.K., 2017. Relationships between land use and stream nutrient concentrations in a highly urbanized tropical region of Brazil: thresholds and riparian zones. Environ. Manage. 60, 30–40.

Tshibanda, J.B., Malumba, A.M., Mpiana, P.T., Mulaji, C.K., Otamonga, J.-P., Poté, J.W., 2021. Influence of watershed on the accumulation of heavy metals in sediments of urban rivers under tropical conditions: Case of N'djili and Lukaya rivers in Kinshasa Democratic Republic of the Congo. Watershed Ecol. Environ. 3, 30–37. https://doi.org/https://doi.org/10.1016/j.wsee.2021.06.001

UNESCO, 2003. Water for People, Water for Life: The United Nations World Water Development Report: Executive Summary. Unesco Pub.

Van der Hoven, C., Ubomba-Jaswa, E., Van der Merwe, B., Loubser, M., Abia, A.L.K., 2017. The impact of various land uses on the microbial and physicochemical quality of surface water bodies in developing countries: Prioritisation of water resources management areas. Environ. nanotechnology, Monit. Manag. 8, 280–289.

Wang, H., Marshall, C.W., Cheng, M., Xu, H., Li, H., Yang, X., Zheng, T., 2017. Changes in land use driven by urbanization impact nitrogen cycling and the microbial community composition in soils. Sci. Rep. 7, 44049. https://doi.org/10.1038/srep44049

Wang, L., Zhang, J., Li, H., Yang, H., Peng, C., Peng, Z., Lu, L., 2018. Shift in the microbial community composition of surface water and sediment along an urban river. Sci. Total Environ. 627, 600–612. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.01.203

Wang, Y., Liu, Y., Huang, Jinming, Wu, T., Huang, Jieyun, 2020a. Analysis and Prevention of Urban River Pollution Analysis and Prevention of Urban River Pollution.

https://doi.org/10.1088/1742-6596/1549/2/022056

Wang, Y., Liu, Y., Huang, Jinming, Wu, T., Huang, Jieyun, 2020b. Analysis and Prevention of Urban River Pollution, in: Journal of Physics: Conference Series. IOP Publishing, p. 22056.

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. Appl. Environ. Microbiol. 78, 8264–8271. https://doi.org/10.1128/AEM.01821-12

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer.

Wilkinson, T.J., Huws, S.A., Edwards, J.E., Kingston-smith, A.H., Siu-ting, K., Hughes, M., Rubino, F., Friedersdorff, M., Creevey, C.J., 2018. CowPI : A Rumen Microbiome Focussed Version of the PICRUSt Functional Inference Software. Front. Microbiol. 9, 1–10. https://doi.org/10.3389/fmicb.2018.01095

Winward, G.P., Avery, L.M., Frazer-williams, R., Pidou, M., Jeffrey, P., Stephenson, T., Jefferson, B., Management, C., Drive, M., Ab, A., 2008. A study of the microbial quality of grey water and an evaluation of treatment technologies for reuse. Ecol. Eng. 2, 187–197. https://doi.org/https://doi.org/10.1016/j.ecoleng.2007.11.001

Xie, Y., Liu, X., Wei, H., Chen, X., Gong, N., Ahmad, S., 2022. Insight into impact of sewage discharge on microbial dynamics and pathogenicity in river ecosystem. Sci. Rep. 1–12. https://doi.org/10.1038/s41598-022-09579-x

Xiong, C., Guo, Z., Chen, S.S., Gao, Q., Kishe, M.A., Shen, Q., 2020. Understanding the pathway of phosphorus metabolism in urban household consumption system: A case study of Dar es Salaam, Tanzania. J. Clean. Prod. 274, 122874.

Xu, J., 2006. Microbial ecology in the age of genomics and metagenomics : concepts , tools , and recent advances. Mol. Ecol. 1713–1731. https://doi.org/10.1111/j.1365-294X.2006.02882.x

Yano, T., Hiromi, K., Hanai, J., Jun, H., Hajime, T., 2013. Stress Tolerance of Methylobacterium Biofilms in Bathrooms. Microbes Environ. 28, 87–95. https://doi.org/10.1264/jsme2.ME12146

Yao, X., Zhang, J., Tian, L., Guo, J., 2017. The effect of heavy metal contamination on the bacterial community structure at Jiaozhou Bay, China. Brazilian J. Microbiol. 48, 71–78. https://doi.org/10.1016/j.bjm.2016.09.007

Ye, L., Zhang, T., 2011. Pathogenic Bacteria in Sewage Treatment Plants as Revealed by 454 Pyrosequencing. Environ. Sci. Technol. 45, 7173–7179. https://doi.org/10.1021/es201045e

Yi, X., Ning, C., Feng, S., Gao, H., Zhao, J., Liao, J., Peng, Y., Zhao, S., Liu, S., 2022. Urbanization-induced environmental changes strongly affect wetland soil bacterial community composition and diversity. Environ. Res. Lett. 17, 14027. https://doi.org/10.1088/1748-9326/ac444f

Young, S., Nayak, B., Sun, S., Badgley, B.D., Rohr, J.R., Harwood, V.J., 2016. Vancomycin-Resistant Enterococci and Bacterial Community Structure following a Sewage Spill into an Aquatic Environment. Appl. Environ. Microbiol. 82, 5653–5660. https://doi.org/10.1128/AEM.01927-16

Zhang, H., Wan, Z., Ding, M., Wang, P., Xu, X., Jiang, Y., 2018. Inherent bacterial community response to multiple heavy metals in sediment from river-lake systems in the Poyang Lake, China. Ecotoxicol. Environ. Saf. 165, 314–324. https://doi.org/https://doi.org/10.1016/j.ecoenv.2018.09.010

Zhang, L., Wang, Z., Cai, H., Lu, W., Li, J., 2021. Long-term agricultural contamination shaped diversity response of sediment microbiome. J. Environ. Sci. 99, 90–99. https://doi.org/https://doi.org/10.1016/j.jes.2020.06.013

Zhang, S.-Y., Tsementzi, D., Hatt, J.K., Bivins, A., Khelurkar, N., Brown, J., Tripathi, S.N., Konstantinidis, K.T., 2019. Intensive allochthonous inputs along the Ganges River and their effect on microbial community composition and dynamics. Environ. Microbiol. 21, 182–196. https://doi.org/https://doi.org/10.1111/1462-2920.14439

Zhang, Y., Ji, G., Wu, T., Qiu, J., 2020. Urbanization significantly impacts the connectivity of soil microbes involved in nitrogen dynamics at a watershed scale. Environ. Pollut. 258, 113708. https://doi.org/https://doi.org/10.1016/j.envpol.2019.113708

Zhao, X., Huang, J., Zhu, X., Chai, J., Ji, X., 2020. Ecological E ff ects of Heavy Metal Pollution

on Soil Microbial Community Structure and Diversity on Both Sides of a River around a Mining Area. Int. J. Environ. Res. Public Health 17, 5680.

Zinabu, E., Kelderman, P., van der Kwast, J., Irvine, K., 2019. Monitoring river water and sediments within a changing Ethiopian catchment to support sustainable development. Environ. Monit. Assess. 191, 455. https://doi.org/10.1007/s10661-019-7545-6

Table S1. Global data sequencing

Catchment	Sample	Total reads	Low-quality sequences	Non-target sequences	Chimeric sequences	Total Valid reads	Read length (bp)	Number of sequences identified (%)	Number of OTUs
Lower	K1D	53137	1575	147	16372	25514	450	31409 (89.6%)	621
	K3D	64521	857	40	21965	46256	448	34837 (83.6%)	971
	N1D	60981	405	162	16337	44077	448	37469 (85.0%)	609
	N2W	50795	337	21187	11099	18172	445	16525 (90.9%)	546
	N4D	46403	664	2	12194	33543	450	29971 (89.4%)	1050
	A3W	66610	518	46	20362	27496	451	404678 (88.6%)	1436
Mid	A4W	75957	471	0	22839	59785	452	48469 (92.1%)	2320
	D5D	52361	348	2055	16978	35043	454	30547 (92.6%)	717
	K1W	44467	536	14	18403	41659	450	32235(87.1%)	1114
	K4D	73861	2058	510	25037	46256	452	41896 (90.6%)	971
	K4W	68504	351	560	17366	50271	452	44111 (87.7%)	962
	K5W	53730	439	231	18740	34320	451	31606 (92.1)	651
	N2D	54338	369	1877	16396	35696	450	30113 (84.4%)	511
	N3D	62089	545	0	21083	40461	452	36439 (90.1%)	621
	N3W	70646	269	6	21420	48951	453	46478 (94.9%)	373
	N4W	83,265	509	503	22,691	59562	454	54380 (91.3%)	655
Upper	PGD	56279	1361	0	27475	28443	448	24956 (87.7%)	1043
	PGW	41521	482	455	12151	28433	448	25628 (90.1%)	895
	A4D	36031	2653	137	5745	52647	280	10956 (39.8%)	561
	BRD	77374	4935	4	12650	74021	320	25579 (42.8%)	697
	BRW	100000	740	0	25239	32980	450	64723 (87.4)	484
	PGW1	28590	1380	838	1536	24836	194	5480 (22.1%)	1205
Total		1321460	21802	28774	384078	896567 (46534)	525 (194-454)		

Table S2. Percent relative abundance of sediment bacterial community from the different catchments zone of two urban streams flowing through Kisumu City

Phylum	Lower catchment						Mid catchment											Upper Catchment					
	N4D	N2W	A3W	A4W	K1D	K3D	K1W	K4D	K4W	K5W	N1D	N2D	N3D	N3W	N4W	A4W	N4W	PGW1	PGW	PGD	BRD	BRW	A4D
AD3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Acidobacteria	0.5	1.3	0.6	0.1	0.8	0.1	0.2	0.8	0.8	1.1	0.5	0.2	0.6	0.0	0.0	0.1	0.0	0.2	0.2	0.0	0.0	0.1	0.1
Actinobacteria	47.0	60.7	51.9	57.9	62.7	62.7	66.5	59.8	50.2	46.7	66.2	69.3	50.1	49.6	42.1	57.9	42.1	7.4	59.0	81.7	3.6	32.7	3.4
Aminicenantes_OP8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Armatimonadetes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BRC1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bacteroidetes	16.0	2.4	4.6	4.2	1.9	5.3	4.1	1.2	2.6	3.8	4.1	0.6	4.5	2.9	2.6	4.2	2.6	1.3	1.2	0.7	1.0	16.4	0.1
Chlamydiae	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlorobi	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chloroflexi	0.1	0.2	0.2	0.5	0.2	0.2	0.0	0.5	0.2	0.2	0.0	0.7	0.1	0.0	0.0	0.5	0.0	1.1	0.2	0.0	0.0	0.6	0.7
Cyanobacteria	0.4	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	20.0	0.0
DQ833500_p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Deinococcus-Thermus	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EU266861_p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Elusimicrobia	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fibrobacteres	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Firmicutes	19.8	15.0	18.1	18.7	13.8	7.7	16.4	25.8	24.6	23.8	7.4	14.7	25.5	27.4	16.6	18.7	16.6	3.8	11.5	8.2	50.6	8.2	42.1
Fusobacteria	1.2	0.0	0.1	0.7	0.0	0.1	0.1	0.2	0.1	0.6	0.3	0.4	0.1	1.5	0.1	0.7	0.1	0.1	0.2	0.0	0.4	1.3	0.0
GQ246397_p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gemmatimonadetes	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.3	0.2	0.1	0.5	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.3	0.0
HE604052_p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lentisphaerae	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nitrospirae	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Parcubacteria_OD1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Peregrinibacteria	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0

Planctomycetes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Proteobacteria	14.6	19.6	24.1	17.9	19.2	23.6	12.6	10.9	20.8	23.1	20.9	14.3	18.0	18.3	38.3	17.9	38.3	75.0	27.7	9.4	41.6	20.3	51.2
SR1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Saccharibacteria_TM7	0.0	0.0	0.2	0.0	0.2	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Spirochaetes	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0
Synergistetes	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
TM6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tenericutes	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Verrucomicrobia	0.1	0.0	0.0	0.0	0.4	0.1	0.0	0.0	0.3	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	10.3	0.0	0.0	2.2	0.0	2.3
WS5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0