Coursework Declaration and Feedback Form

The Student should complete and sign this part

Student Number: 2531311X	Student Name: Wei Xiao
Programme of Study (e.g., MSc in Electronics and Electrical Engineering):	
Course Code: ENG5059P	Course Name: MSc Project
Name of **First** Supervisor: Dr Umer Ijaz	Name of **Second** Supervisor: Carlos Domingo-Felezg
Title of Project: Exploring Microbial Correlations and Diversity in Pit Latrines	
Declaration of Originality and Submission Information	
I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements Signed (Student): Wei Xiao	
Date of Submission: 14th August 2023	

Feedback from Lecturer to Student to be completed by Lecturer or Demonstrator	
Grade Awarded: Feedback (as appropriate to the coursework which was assessed):	
Lecturer/Demonstrator:	Date returned to the Teaching Office:

Exploring Microbial Correlations and Diversity in Pit Latrines

Student: Wei Xiao

Student ID: 2531311X

Supervisor: Dr Umer Ijaz

Co-supervisor: Dr Carlos Domingo-Felezg

August 14th, 2023.

A thesis submitted in partial fulfilment of the requirements for the degree of MASTER OF

SCIENCE IN COMPUTER SYSTEMS ENGINEERING

ACKNOWLEDGEMENT

I am deeply grateful for the patient guidance of Dr. Umer Ijaz, my mentor. I extend my heartfelt appreciation for introducing me to the captivating realm of microbiology, a field previously unexplored by me. He has consistently followed our project's progress and generously assisted us in overcoming any research challenges.Furthermore, I sincerely thank Uzma Uzma for her insightful elaborations on network inference, and Kelly J. Stewart for her enthusiastic support during challenging phases of my paper. I also wish to express my gratitude to Aqsa Ameer for her valuable suggestions on my poster presentation.

In addition, I am thankful to my fellow group members; each interaction with them has deepened my understanding of the course materials.Embarking on a journey in a completely new field is an immensely challenging task. I consider myself fortunate to have received strong support from these outstanding individuals, who have enriched my learning journey in multiple ways. My year at the University of Glasgow has been a transformative experience. Starting in a non-native language environment was initially challenging, but the continuous presence of helpful individuals, whether in daily life or academic pursuits, has made my experience vibrant and fulfilling.

As I bid farewell to Glasgow, a sense of nostalgia fills me. I believe this cherished journey will continue to accompany me. If given the opportunity, I would revisit the path I've traveled, witness the sights I've seen, and relish the same warmth that different people have bestowed upon me.

ABSTRACT

Although improving personal hygiene, access to clean drinking water, and sanitation facilities can prevent 2.4 million deaths (Bartram and Cairncross, 2010), economic constraints lead to the majority of the world's poor having no choice but to rely on pit latrines, estimated to be used by 1.77 billion people daily (Graham & Polizzotto, 2013). Pit latrines are a low-cost method that separates human excreta from people, reducing the spread of diseases by reducing the amount of human feces in the environment. However, the main issue with these latrines is that they eventually fill up and require emptying or replacement, causing long-term economic burdens. In theory, increasing the decomposition rate of microbes inside the pit can extend its lifespan, delaying the need for emptying. Therefore, studying the coexistence and diversity of microbes inside the pit can help us understand the factors influencing decomposition rates.

The focus of this study is to analyze microbial abundance data collected from different sampling depths within pit latrines. We employed network inference techniques to explore potential microbial interactions and construct a microbial interaction network. By analyzing co-occurrence patterns and correlations, we identified influential relationships between key microbes and other microorganisms within the microbial community. Additionally, we used the zeta distribution, a probability distribution that characterizes species abundance distributions, to assess microbial diversity. By fitting the zeta distribution to the microbial abundance data, we gained deeper insights into the diversity and relative abundance of microbial taxa within the pit latrines.

In this study, through network inference, we found that microbial distributions in pit latrine samples from Vietnam exhibited long-tail characteristics, while those from Tanzania did not show significant structural features. Moreover, the lifestyle habits of local residents influenced the population and influence of microbes within the latrines. For example, in Vietnam, the use of lime for cleaning led to an increase in the proportion of alkaliphilic bacteria in the samples. In the microbial network graph from Vietnam, Methylobacter demonstrated the highest influence due to its ability to create an acidic environment, forming mutualistic relationships with certain bacteria. In Tanzania, Paludibacteraceae H1 exhibited the highest influence because it participates in preliminary hydrolysis and acid production, providing metabolic substrates to other microbes. These findings indicate that human lifestyle significantly alters the microbial community distribution within pit latrines.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Although improvements in household water supply, environmental sanitation, and personal hygiene (WASH) have been made recently, roughly half of the world's population, or roughly 4.5 billion people, still do not have access to better sanitation facilities(Soares Magalhães et al., 2011). A large portion of this population lives in low-income regions in Africa, Asia, Latin America, the Caribbean, and the Pacific, as well as the Pacific Islands (Willis Gwenziet al. ,2023). Out of the total population, an estimated 670 million individuals persist in engaging in open defecation as a consequence of lacking access to sanitary facilities. According to Torondel et al. (2016), impoverished urban and rural individuals often have limited access to inexpensive sanitation facilities. Consequently, the most viable alternative for enhancing sanitation conditions in these communities is the implementation of on-site sanitation systems, such as pit latrines. Pit latrines are widely recognized as the most suitable sanitation option for low-income settings. They typically consist of an excavated pit equipped with a platform covered by a slab with a drop hole.

Additionally, a superstructure is provided to ensure user privacy. According to Willis Gwenziet al. (2023), the typical expenditure for constructing rudimentary pit latrines in underdeveloped nations is $25 to $60. These facilities play a crucial role in mitigating the transmission of infectious and parasitic diseases by effectively isolating fecal matter from the surrounding environment. Additionally, they offer individuals the benefits of enhanced privacy, convenience, and comfort. According to Bartram and Cairncross (2010), the implementation of measures such as enhancing personal hygiene practices, ensuring the availability of clean drinking water, and providing adequate sanitation facilities has the potential to avert around 2.4 million deaths on an annual basis.

A significant obstacle associated with the utilization of pit latrines is the eventual depletion of the pit's capacity, necessitating the need for either emptying or replacement (Ijaz UZ, 2022). The financial implications associated with emptying or replacing pit latrines can be substantial, resulting in a persistent economic burden over an extended period. In order to investigate this matter, the present study seeks to utilize 16S rRNA sequencing, zeta diversity analysis, and network inference methodologies to examine the microbial communities derived from toilet samples collected in Vietnam and Tanzania. Valuable insights can be derived for developing sustainable sanitation facilities in places with limited resources by identifying diversity and co-occurrence patterns.

1.2 Outline

Chapter 1 discusses the role of pit latrines in curbing disease transmission in developing countries while facing the challenge of emptying or rebuilding once filled. This chapter mentions the objectives and significance of the project, highlights the necessity of the research, and refers to completed studies in related fields.

Chapter 2 covers the sampling areas and processes, including toilet selection, fecal sample collection and analysis, the principles and applications of zeta analysis and network inference, and how these analytical methods are combined to interpret the results.

Chapter 3 provides an interpretation of the results obtained from the methods described in Chapter 2.

Chapter 4 further interprets the results from Chapter 3 and summarizes the conclusions.

Chapter 5 discusses potential future research and presents reasonable recommendations in the field.

1.3 Microbial Distribution in Pit Latrines - Relevant Research

1.3.1 Introduction to Pit Latrines

In many developing nations, pit latrines are a typical type of toilet used to collect human waste and store it in underground pits to lower the risk of disease transmission via open defecation. According to a report by UNICEF (2022), pit latrines are widely utilized as a sanitation solution on a global scale, serving as a primary facility for about 1.77 billion individuals.

As seen in Figure 1, the primary constituents of a pit latrine comprise the subterranean pit, a concrete platform or base, and a structure serving as a shelter, such as an improvised dwelling or a dedicated toilet enclosure. Guidelines provided by UN-Habitat state that "Pit latrine pits are typically at least three meters deep and one meter wide; the size of the small hole in the floor should not exceed 25 centimeters to prevent accidents" (UN-Habitat, 2010). The shelter offers a sense of seclusion and security for individuals, safeguarding them from inclement weather circumstances.

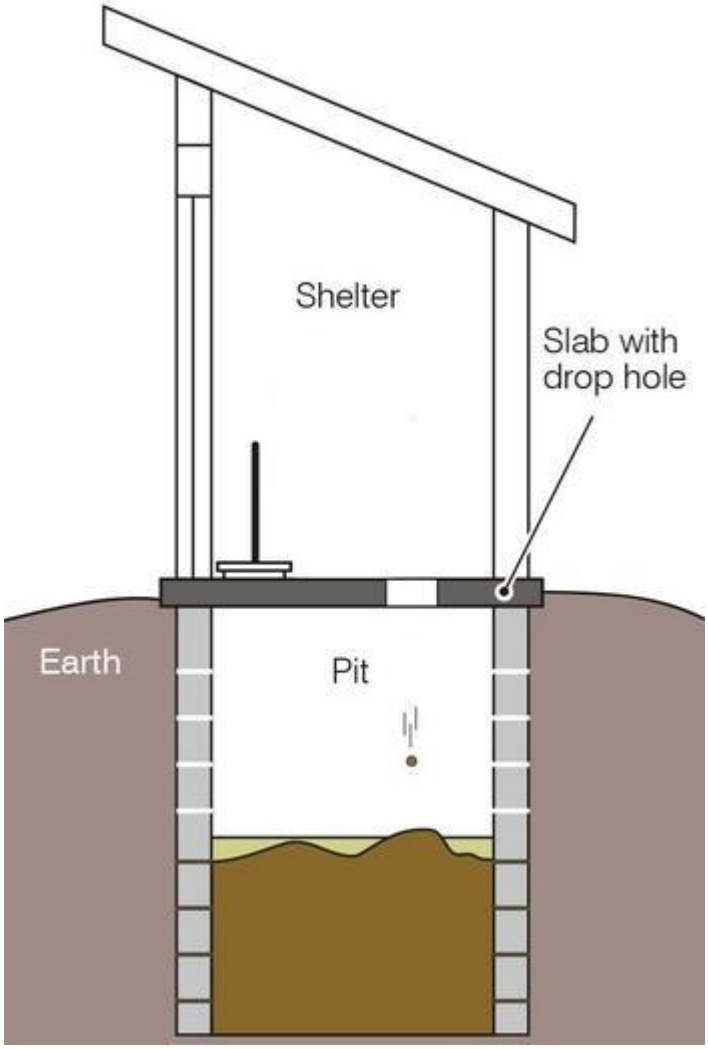


Figure 1:Components of a classic pit latrine from WEDC publication in 2014

Pit latrines can be classified into two main types: dry and flush. The World Health Organisation (WHO, 2019) states in its report that dry-pit latrines do not rely on water for flushing. However, flush-pit latrines incorporate water seals and utilize minimal water for flushing. Water seals play a crucial role in mitigating odors, minimizing the breeding of flies, and reducing the potential for contamination. Pit latrines may successfully limit the amount of human waste in the environment when built and maintained correctly, which reduces the spread of diseases brought on by open defecation.

Nevertheless, pit latrines confront difficulties in the process of pit emptying and the management of faecal sludge. Based on the estimations provided by the World Health Organisation (WHO, 2019), it is indicated that there is a global requirement for the emptying and administration of roughly 450 million pit latrines. Ensuring the safe emptying of pits and effective handling of faecal sludge is crucial in mitigating water pollution and safeguarding public health. The "Clean India Mission" in India aims to end open defecation and improve sanitation infrastructure, which includes pit latrines, in order to solve this issue. Other efforts and movements have also been implemented to promote pit latrine construction and use.

Pit latrines generally offer fundamental sanitation services, meeting the communities' basic hygienic needs without modern restrooms. However, initiatives are required to ensure secure and long-lasting pit emptying and faecal sludge management, as well as to advance public hygiene and health practices and support social development.

1.3.2 The Importance of Pit Latrines

Implementing pit latrines plays a pivotal role in mitigating the prevailing global sanitation dilemma and guaranteeing the provision of fundamental sanitation amenities to billions of individuals around the globe. Pit latrines serve as a crucial intervention, particularly in low-income nations where insufficient sanitary infrastructure presents significant health hazards. Lindholm and Larsson (2022) argue that the implementation of improved sanitation infrastructure, such as pit latrines, plays a vital role in both curbing the transmission of waterborne infections and minimizing the consequences of diarrheal ailments.

When infrastructure and resources are scarce, pit latrines are a viable and affordable option for safe excreta disposal (Mara et al., 2010). This makes it challenging to implement more sophisticated sanitation systems. To reduce the pollution of water sources and the transmission of diseases, a basic degree of sanitation is implemented through the facilitation of collection and treatment of human waste, as well as the prevention of open defecation.

In addition, the implementation of pit latrines serves a key role in preserving the quality of groundwater. The issue of groundwater contamination resulting from insufficient sanitation practices, such as the utilization of pit latrines, is a matter of great significance, particularly in regions where groundwater serves as the primary drinking water source. Adequate separation between pit latrines and water sources can be achieved through proper construction and appropriate siting, reducing the potential for groundwater pollution (Caldwell & Parr, 1937).

Communities can gain from improved sanitation conditions, less waterborne illness transmission, and protect groundwater resources by encouraging the widespread use of pit latrines.

In general, implementing pit latrines is essential for effectively mitigating the sanitary predicaments a substantial global population encounters. These rudimentary and affordable sanitation facilities are crucial in advancing public health, mitigating the spread of waterborne diseases, and protecting groundwater reservoirs.

1.3.3 Obstacles and Difficulties

One significant issue with full pit latrines is that they eventually run out of room for excreta, necessitating emptying or replacement. This topic covers the management of diseases and hazardous substances included in human waste, as well as the purchase of appropriate equipment and technology for the transportation and treatment of such waste materials.

Fecal matter has the potential for carrying a range of diseases, including bacterial, viral, and parasitic pathogens (Strande et al., 2014). If these microorganisms are not appropriately managed, they possess the capacity to facilitate the transmission of diseases, hence bringing risks to both the environment and public health (World Health Organization, 2014).

In addition, it is important to recognize that full-pit latrines involve a diverse array of potentially hazardous substances, including nitrogen compounds, phosphates, heavy metals, organic compounds, and pharmaceutical residues (Drechsel et al., 2015). The excessive presence of nitrogen and phosphates can give rise to pollution in groundwater and water bodies, hence causing subsequent disruptions to the biological balance within aquatic ecosystems. The presence of heavy metals in the environment can be attributed to the release of wastewater and waste materials, leading to negative consequences for both the ecosystem and various organisms occupying it(Rohwer, 2017). The potential impact of organic compounds and pharmaceutical residues on water bodies extends to both the environment and human well-being.

Pit latrines must be emptied or replaced with new ones, which necessitates the use of the proper tools and technology, including vehicles, devices, and containers for collecting and transporting excreta, as well as trained persons and equipment for emptying and treating them (Strande et al., 2014). However, in contexts with limited resources, purchasing and operating such technology may provide difficulties, particularly for underdeveloped and distant places, raising the price of pit latrines.
Hence, extending the operational lifespan of pit latrines can yield substantial cost reductions.

1.3.4 Literature Review on Microbial Profiles of Pit Latrines

Researchers in the field of pit latrine microbiology have achieved significant achievements in the investigation of microbial symbiosis and variety. Pit latrines support various microbial communities, encompassing microorganisms such as bacteria, fungus, viruses, and parasites. Communities of microbes engage in intricate interactions and perform pivotal activities, including the decomposition of organic waste and the transformation of faecal matter. Peiffer et al. (2019) observed that the presence of various pit latrine designs and environmental conditions might result in divergent microbial communities, highlighting the impact of environmental factors on the composition and functionality of microorganisms. Researchers have investigated the ecological functions of bacteria in filled pit latrines by examining their metabolic activities and the expression of functional genes. This research has offered valuable insights into optimizing waste treatment and resource recovery. In general, scholarly investigations about the microbiological symbiosis and diversity within filled pit latrines offer valuable scientific backing for comprehending the microbial ecology in such latrines. Furthermore, these studies enhance the efficiency of waste treatment processes while concurrently ensuring the protection of both the environment and public health.

1.3.5 Applications of Pit Latrine Contents

Pit latrines have a notable impact across multiple domains, particularly in contexts characterized by low resources and in underdeveloped nations. Pit latrines in rural regions serve as essential sanitation infrastructure, contributing to enhanced living circumstances and improved sanitation standards within the local environment. Pit latrines are crucial in meeting basic hygienic requirements in geographically and environmentally diverse isolated areas. Pit latrines play a crucial role in disaster relief and refugee camps by enabling swift deployment, efficiently managing excreta, and mitigating the spread of infectious diseases (Sabogal, 2014). Furthermore, inside the urban slums of poor nations, pit latrines are crucial in enhancing sanitation conditions and offering fundamental sanitary amenities. In various contexts, such as rural areas, remote regions, emergency relief, refugee camps, and urban slums in developing nations, pit latrines

fulfill significant functions by offering individuals fundamental sanitary amenities and enhancing their overall living conditions.

1.3.6 Pit Latrine Processing

In order to gain a comprehensive understanding of microbial activities within pit latrines, it is essential first to establish a preliminary understanding of the constituent composition and degradation processes occurring within these latrines. The initial composition of a pit latrine predominantly comprises recently excreted human faeces, encompassing excreta, urine, and sanitation products. Fresh excrement contains organic components that are readily biodegradable. In specific cases, additional waste products are deposited into the pit, resulting in heterogeneous mixing. The materials within a pit latrine exhibit stratification, forming separate layers according to their respective positions. These layers experience different levels of deterioration and decomposition mechanisms. Fresh faeces make up most of the top layer, along with other materials still subject to microbial deterioration.

The second layer comprises the top aerobic stratum, where aerobic conditions facilitate the rapid hydrolysis of organic materials. The third layer experiences a decelerated anaerobic digestion process due to the reduction of oxygen caused by the presence of covering materials. This depletion leads to the decomposition of complex organic compounds into simpler ones. The fourth layer, located at the lowest level, ceases to facilitate organic decomposition, resulting in the materials reaching a state of stability. Microbial involvement is a crucial factor in the biodegradation processes inside these layers. A substantial amount of oxygen within the top aerobic layer enables the efficient aerobic degradation of organic substances into smaller molecular molecules. In the subterranean strata, when deprived of oxygen, bacteria engage in anaerobic degradation of organic matter, breaking complex compounds into simpler molecular constituents. As the depth increases, the degradation rates steadily decrease, increasing the stability of organic matter. The upper layer has the most rapid degradation, while the bottom layer demonstrates the slowest degradation rate. Various factors, including the patterns of water infiltration, user behaviours, and oxygen availability, influence the degradation processes occurring in pit latrines. Various variables can contribute to the variability in material qualities observed across different latrines. The characteristics and stability of latrine contents are impacted by the decomposition and transformation of organic matter caused by microbially mediated aerobic and anaerobic degradation processes within pit latrines (Torondel et al., 2016).

1.4 Aims/Objectives of The Project

1. The objectives of the project are as follows:
2. To identify the microbial communities present in pit latrines.
3. To analyze the diversity of microbes present in pit latrines.
4. To uncover the symbiotic relationships among the microorganisms.
5. To quantify the correlations between different bacterial species.
6. Discover the factors that influence the distribution of microorganisms.

CHAPTERCHAPTER 2

METHODOLOGY

2.1 Study Area and Sample Selection

For this investigation, 22 pit latrines were chosen as samples from the Hoang Tay and Nhat Tan communities close to Hanoi, Vietnam. Eight pit latrines were also selected as samples from the Sululu and Signali villages near Ifakara, Morogoro area, Tanzania. In Vietnam, the pit latrines are constructed in an elevated manner, intended explicitly for faecal collection to be utilised in agricultural practises. Typically, these latrines possess limited pit capacities and depths that do not surpass 1 metre. According to Torondel et al. (2016), users of Vietnamese latrines frequently sprinkle ash and lime, which could affect pH levels and cause other environmental changes. The mentioned practise can impact the long-term viability of bacterial populations in the excrement that accumulates within toilet facilities. In the context of Tanzania, latrines were selected in the Sululu and Signali villages located near Ifakara. These villages are inhabited by individuals primarily involved in subsistence farming and follow a predominantly vegetarian diet pattern. The latrines in this particular region are often built within the soil, with an approximate depth of 2 metres.

The objective of our study is to perform 16S rRNA gene DNA sequence analysis on the microbial communities present in the latrines of both nations. The microbial composition of the samples is analysed both qualitatively and quantitatively to investigate the extent of microbial variety present in the latrines and uncover any symbiotic interactions that may exist among them. In order to attain these aims, we took into account other attributes of the samples, such as the existence or lack of a roof structure above the latrine, the composition of the walls (grass or bricks), the presence or absence of a liner within the pit, and the squatting slab material (soil or cement/brick). Furthermore, an examination was conducted on the distribution patterns of individuals utilising latrines, their methods of faeces disposal (including the implementation of urine separation), utilisation of toilet paper and water, as well as their food habits.

2.2 Methodology for Sample Collection and Analysis

In the soil sampling process, conventional soil augers or soil augers outfitted with sterile plastic containers were employed to conduct layer-by-layer sampling. A total of 200 grams of substance was gathered from individual pit latrines, with samples taken at 20 centimeters spanning from the uppermost to the lowermost regions. The samples that were gathered were carefully stored in containers that had been sterilized. Specialized samplers were employed when the liquid layer surpassed a measurement of 20 centimeters. The sampling apparatus comprised sampling tubes equipped with a gate featuring spring-loaded flaps. The samples were obtained using sterile containers and promptly transferred to a refrigerated container for subsequent analysis on the collection day.
The environmental parameters, namely pH, temperature, total and soluble chemical oxygen demand (CODt and CODs), volatile fatty acids (VFAs), total solids (TS), volatile solids (VS), ammonia, total phosphate, carbohydrate, and protein, were assessed according to the methodology described in Torondel et al. (2016).

Once the collection process had been concluded, the extraction of bacterial DNA from the collected samples was performed. This extraction specifically targeted the microbial communities in pit latrines from two distinct countries. Following this, polymerase chain reaction (PCR) amplification was conducted on the DNA extracted from the bacteria. The primers used in this process were designed to specifically target sections 3 to 5, which are variable portions of the 16S ribosomal RNA (rRNA) gene. Distinctive DNA barcoded primers were employed to label and differentiate each sample. Following that, the DNA samples were tested with high-throughput sequencing technology, specifically 454 pyrosequencing, in order to generate significant bacterial 16S rRNA sequence data. The AmpliconNoise workflow was used for the processing of the sequencing data. The process encompassed numerous phases, including the removal of signals of low quality, the filtration and trimming of sequence data, and the identification and elimination of noise and chimeric sequences. The processed sequences were assigned to individual samples based on their respective DNA barcodes. The application of unsupervised clustering was utilized to classify the sequence data, with the aim of grouping bacterial sequences that displayed similarity into operational taxonomic units (OTUs). The clustering procedure utilized a sequence similarity threshold of 3%, while sequences exhibiting a similarity over 97% were assigned to the same operational taxonomic unit (OTU). The results of OTU clustering were employed to generate an OTU table, which gathered the abundance data of each unique OTU in every sample. The structure of this table can be described as a matrix, where the rows represent samples, the columns correspond to operational taxonomic units (OTUs), and the values indicate the relative abundance of each OTU within its corresponding sample. Therefore, the OTU table reveals significant insights into the composition and organization of the microbial population.

Figure 2: The standard auger used to collect the faecal samples from the pit latrines in Tanzania and Vietnam (Torondel, et al., 2016).

2.3.1 Zeta Diversity is a measure used in ecological studies to quantify species diversity within a habitat or community. It is a metric that

Zeta diversity refers to a quantitative measure utilized to evaluate the extent of dissimilarity or difference in species composition across different populations or regions.
The method integrates data on the presence or absence of species and their relative abundance to evaluate the variance in composition across several places. As proposed by Jost (2007), the concept of zeta diversity extends the conventional metrics of alpha and beta diversity by considering both species turnover and nestedness. Alpha diversity measurements typically center on quantifying the number of species and the diversity inside a given sample. Conversely, beta diversity metrics emphasize discerning dissimilarities and similarities among many samples. Nevertheless, these metrics must consider the specific number of species shared and how they are distributed among various samples. Consequently, representing diversity through solely alpha and beta diversity becomes inadequate when handling three or more samples. On the other hand, zeta diversity considers the number of species shared across numerous samples. As mentioned earlier, the statement offers a more thorough and integrated depiction of the distribution of diversity. The index is between 0 and 1, where higher values signify increased group dissimilarity. As the diagram shows, alpha diversity is denoted as $\zeta 1$, beta diversity as $\zeta 2$, and zeta diversity as $\zeta 3$. The intersection between two given samples A and B, is represented as $A \cap B$ ($\zeta 2$), while the overlap among samples A, B, and C in the diagram represents zeta diversity ($\zeta 3$).

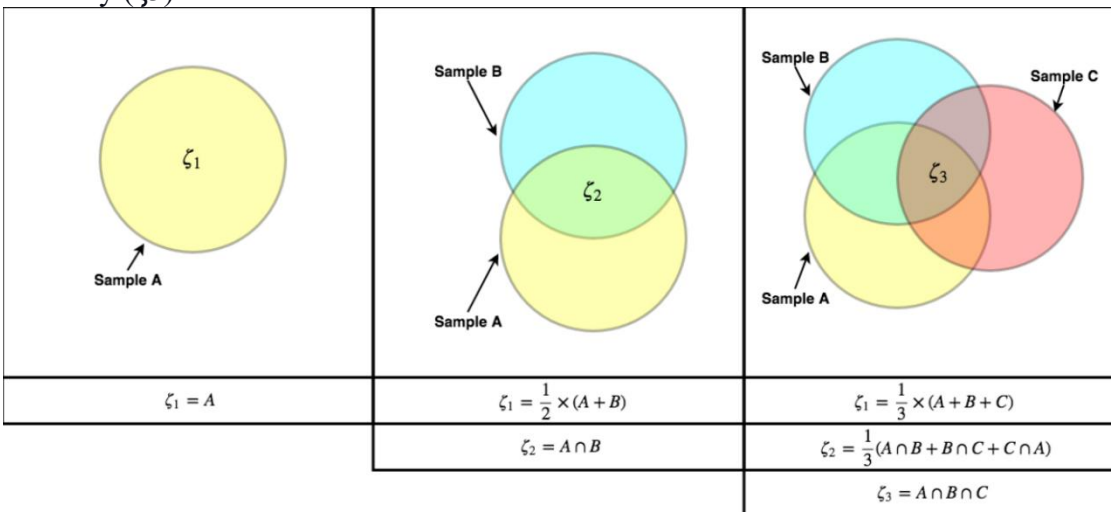


Figure 3:Comparison of α -diversity, β -diversity and ζ -diversity(Zeta Diversity - Wikipedia, 2018)

2.3.1 Analysis of Zeta Diversity

Zeta diversity analysis was employed to examine the distribution of microbiological variety in samples collected from pit latrines in both Vietnam and Tanzania to identify and compare their distinguishing characteristics and shared features. The analysis was conducted with the "zetadiv" package.

The Zeta Diversity Decline metric is employed to examine the co-occurrence of species and measures the extent of diversity variation concerning species co-occurrence. As the number of samples rises, there is typically a steady decrease in Zeta diversity loss. This phenomenon can be attributed to a more significant number of samples resulting in a more excellent catch of various

species, hence causing an augmentation in species co-occurrence. At lower levels of species co-occurrence, the augmentation of species co-occurrence exerts a more pronounced influence on reducing Zeta diversity. Consequently, the introduction of extra samples has the potential to result in a more substantial reduction in Zeta diversity. As the sample size increases, the impact of increased species co-occurrence on the decline of Zeta diversity diminishes, as a significant proportion of the species has already been taken into consideration. This fact indicates that as the number of samples increases, the relative contribution of less frequent species to the overall community composition decreases.
As a result, in situations where species co-occurrence is more common, the impact of introducing additional samples on the reduction of Zeta diversity is less pronounced. The result of this phenomenon is a progressive reduction or stabilizing of the diminishing trend. McGeoch et al. (2019) argue that the decline in Zeta diversity provides valuable information regarding the key species present in the community.

The decrease in zeta diversity is determined by the ratio of zeta values, which are constrained within the numerical interval of 0 to 1. As zeta values approach critical points, their ability to differentiate between different levels of diversity may diminish (Bennett & Gilbert, 2015). In order to tackle this issue, the fall in the Zeta diversity ratio is employed for identifying purposes. Calculating the Zeta diversity decline ratio involves comparing Zeta diversity decrease values between consecutive orders. To illustrate, the Zeta diversity decline values for orders 2, 3, 4, and 5 will be calculated. Subsequently, the ratio of Zeta diversity decline will be determined by dividing the Zeta diversity decline at order three by the Zeta diversity decline at order two and by dividing the Zeta diversity decline at order four by the Zeta diversity decline at order 3, and so forth. By examining alterations in this ratio, one can gain insight into the fluctuations in interactions between species co-occurrence. A larger ratio signifies an increased involvement of a specific order in species co-occurrence, implying the existence of pivotal species or environmental factors that hold significant influence at that order. A rising ratio implies a greater decrease in species co-occurrence, indicating more significant alterations in species composition across samples and a less stable co-occurrence pattern. In contrast, a declining ratio indicates a reduced pace of variation in species co-occurrence, indicating greater stability at higher levels.

2.3.2 Network Inference

Network inference is applied to comprehend the interactions and relationships among the bacteria in pit latrines and how they affect the environment and operation of the latrine. Microbial networks can be built using the sample data gathered, where microbial species are represented as nodes and their interactions as edges. Examining the network architecture and computing various centrality measures can determine the most critical nodes. This network approach can reveal symbiotic, antagonistic, and cooperative connections between bacteria. The primary techniques employed for network inference encompass IVI (Iterative et al.), hubness score, and spreading score.

1. Hubness Score: The Hubness Score is a comprehensive metric derived from combining two centrality indices, namely Degree Centrality and Local H Index. Degree Centrality is a statistic used to measure the number of direct connections that a node has with other nodes, which is referred to as its degree. In contrast, the Local H Index is a centrality index that is

calculated by taking into account the degrees of the adjacent nodes of a specified node(Abbas Salavaty al., 2020).
To calculate the Hubness score, the Degree Centrality and Local H Index data are first normalized to address the potential influence of their different scales on the results. The next procedure is the computation of a comprehensive score that signifies the impact of the node on its immediate environment. The calculation of this score involves a combination of the normalized values of Degree Centrality and Local H Index.

The Hubness score provides a more comprehensive assessment of the centrality and importance of nodes within a network. Through the consideration of both a node's direct connections and those of its neighboring nodes, a more precise portrayal of the node's influence on the entire network is attained(Lovell et al., 2015).

$$\text{Hubness Score} = \text{Degree Centrality} + \text{Local H Index} \quad (1)$$

2. Spreading Score: The Spreading Score is a widely used centrality measure in network analysis that is utilized to estimate a node's ability to distribute information throughout the network. The concept of identifying important and capable nodes for information transmission inside a network was first introduced by Chen et al. in 2012.
Information dissemination inside a network refers to the method through which knowledge or influence is spread across the connected nodes. For instance, information can flow through links between friends in a social network. In contrast, in microbiology, interactions between bacteria and other microbes can affect their survival and position within the ecosystem. These interactions can be symbiosis, antagonism, competition, and other interactions.

The computation of the Spreading score entails the simulation of information dissemination within the network, considering the node's connections, namely its neighbouring nodes, and the corresponding connection weights. A node may have an advantage in information dissemination if it is connected to numerous other nodes, and these connections have greater weights. The spreading process can be replicated by utilizing random walks or alternative propagation models.

When determining the Spreading score, the inclusion of probabilities about the dissemination of information to a node within the network is taken into account during the random selection of a node as the initial point. The dissemination process considers the connectivity of the node, specifically its neighbouring nodes, as well as the magnitudes of the connections. This procedure facilitates the assessment of the node's capacity for dissemination. The mathematical expression used to compute the Spreading score is as follows:

$$Spreading_{score_i} = (NC'_i + CR'_i)(BC'_i + CI'_i) \quad (2)$$

Where NC,CR,BC,CI are range normalized neighborhood connectivity, ClusterRank, betweenness centrality, and collective influence of node i, respectively(Abbas Salavaty al., 2020).

3.The influence and vulnerability of nodes in a network are evaluated using the comprehensive metric known as the IVI (Influence-Vulnerability Index). A node's potential to

impact the entire network is represented by influence, while the vulnerability indicates its susceptibility to outside disruptions or attacks.
Spreading and Hubness ratings, previously described, show different aspects of each node. The Hubness score measures a node's performance in its immediate surroundings, but the Spreading score indicates its growth potential. The influence of a node within the overall network is directly proportional to the product of its Spreading and Hubness ratings(Abbas Salavaty al., 2020).

The Index of Vertical Integration (IVI) is determined by the mathematical process of integrating the Spreading and Hubness scores. This integration is performed using a Multiplication function, which allows for the capturing of the combined and mutually reinforcing impact of these scores. By calculating the IVI, nodes inside the network can be assigned a thorough rating, facilitating the discovery of nodes with more significant influence and lower susceptibility. As mentioned above, the nodes exert substantial influence on the dynamics and resilience of the network in response to any disturbances or perturbations.

$$IVI_i = (H\text{ubness}_{score_i})(Spreading_{score_i}) \quad (3)$$

$$IVI_i = (DC_i' + LH'_{index_i})((NC_i' + CR_i')(BC_i' + CI_i')) \quad (4)$$

Where A denote the adjacency matrix of the network ,and each element Aij represents the connection between nodes i and j. Specifically, Aij equals 1 if nodes i and j are connected, and 0 otherwise. The operators K and N are utilized in measuring the degree and set of first-order neighbors of a node, correspondingly(Abbas Salavaty al., 2020). The operator $\mathscr{H}$ is utilized to compute the H index of node i by considering the degree of its adjacent nodes. The term f(ci) incorporates the influence of i's local clustering. Let Smn denote the quantity of shortest paths connecting nodes m and n, whereas Smn(i) represents the quantity of shortest pathways connecting nodes m and n that traverse through node i(Abbas Salavaty al., 2020). The notation $\delta B(i,\ell)$ denotes the collection of nodes that are located at a distance ℓ from the node i(Abbas Salavaty al., 2020).
The Min-Max feature scaling method is employed to normalize the centrality metrics, ensuring they are within a consistent range. This normalization process eliminates any potential bias caused by varying weights, while maintaining the relative weight ratios of the measures.

4.Phi：Inside the realm of life sciences, numerous measurement techniques solely offer the comparative proportions of distinct constituents inside a given specimen (Lovell et al., 2014). Caution must be exercised when analyzing differential expression due to the compositional character of the relative data. Conventional statistical methodologies, such as correlation analysis, are not suitable for the examination of relative data. The utilization of correlation as a method for examining relative abundances has the potential to yield conclusions that are contrary to those derived from absolute abundances. Furthermore, the correlation values themselves may vary when other components are incorporated into the analysis, as noted by Lovell et al. (2014). In order to effectively tackle the concerns associated with relative data, this study employs a unique statistical measure, ϕ (Phi), as a means to articulate the degree of proportionality between two variables. The formula for the statistic phi can be represented in practical terms as follows:

$$\phi(\log x, \log y) = \frac{\mathrm{var}(\log \frac{x}{y})}{\mathrm{var}(\log x)} \qquad (5)$$

Can also be expressed as:

$$\phi(\log x, \log y) = 1 + \beta^2 - 2\beta |r| \qquad (6)$$

Here, β represents the slope estimated by the standardized major axis (SMA), and r is the correlation coefficient between the logratio values of the two variables. The value of φ ranges from 0 to positive infinity, and the closer it is to 0, the higher the proportionality between x and y, indicating a more linear relationship. When φ equals 0, it means x and y are perfectly proportional, i.e., there is a direct linear relationship between their values. On the other hand, when φ approaches positive infinity, it implies a complete inverse relationship between x and y, i.e., they are reciprocally related.

2.3.3 The present study focuses on two regression modeling techniques: power law and exponential regression.

Power Law Regression and Exponential Regression are two mathematical models employed for fitting data that correspondingly display power-law growth or decay and exponential growth or decay patterns. These two regression curves can be used to fit the zeta diversity decline curve and find the best-fitting growth or decay curve in studying microbial zeta distributions. Regression can also be used to assess how scale dependency in a community's species occurrence, which is correlated with zeta diversity ratio, operates (Hui & McGeoch, 2014). The Power Law Regression reflects scale dependence in species composition, with a higher likelihood of retaining more widespread species at finer scales (Hui & McGeoch, 2008; McGlinn & Hurlbert, 2012). The Exponential Regression reflects scale independence in species retention across different locations.

The mathematical expression representing Exponential Regression can be defined as y is the product of a and the exponential function of b multiplied by x.
The mathematical expression representing Power Law Regression can be defined as y equals a multiplied by x raised to the power of b.

In this context, the variable y denotes the magnitude of the zeta diversity drop, x represents the ordinal position of zeta diversity, and a and b are the regression model coefficients. The 'exp' function denotes the exponential function. The fitting procedure aims to identify suitable values for a and b that reduce the fitting error between the exponential regression or power law regression curve and the zeta diversity drop curve. Researchers can obtain estimated values for the regression coefficients a and b by employing regression curve fitting. Within the current context, the coefficient denoted as "a" represents the initial value of the zeta diversity drop curve. In contrast, the coefficient denoted as "b" characterizes the slope or incline of this curve. The fitting results produced can be employed to interpret the patterns and characteristics of the zeta diversity drop curve. This entails assessing whether the curve exhibits an exponential growth or decay pattern, a power-law growth or decay curve, and how it changes over different zeta diversity orders.

The outcomes of the fitting process yield significant insights into the dynamics of microbial communities, illustrating the changes in community structure across diverse geographical and temporal scales. The findings presented in this study provide a greater understanding of how the

migration capacities of species and the consequences of isolation influence the general composition of the community. Furthermore, modern discoveries have enabled the ability to predict and simulate the fluctuating modifications within microbial communities through exponential or power-law growth or decay curves.

Suppose the Power Law Regression model better matches the zeta diversity drop curve. In that case, it indicates that the microbial community in the pit latrine has attributes consistent with a distribution with a large tail. This phenomenon indicates that as the depth of a given system rises, the rate at which zeta diversity decreases gradually decelerates. The observed trend suggests a limited number of dominating species or functional groups that consistently occur at various depths within the samples. In contrast, other species or functional groups exhibit a steady decline. The long-tailed distribution is frequently observed in ecosystems characterized by species possessing robust migratory capabilities. This suggests that microorganisms exhibit a heightened migration rate between samples from varying depths, permitting frequent interchange and movement.
Moreover, when species distribution in an ecological network adheres to a power-law regression distribution, it indicates that a substantial quantity of uncommon species exists. Within this particular distribution, a limited number of species exhibit a significant prevalence within the overall population, while a relatively low number of individuals characterizes most species. The distribution pattern in question is widely recognized in academic literature as the "long-tail distribution" or "Preston curve."

Certain species exhibit elevated population densities in the ecological network characterized by a power-law regression distribution, whereas others demonstrate exceptional rarity. This phenomenon can be linked to various species' differential availability of resources and habitats, influencing their population sizes. Some species may have greater access to resources and habitats, enabling them to support more significant populations. Conversely, other species may have disadvantages in resource competition, leading to smaller population sizes. The presence of uncommon species plays a vital role in the ecological functioning of an ecosystem. These uncommon species frequently assume significant ecological functions, such as occupying specialized ecological niches, facilitating seed dispersal, or exerting influence over crucial ecological processes. Simultaneously, the depletion of uncommon species might have detrimental effects on the stability of the ecosystem. Hence, the manifestation of a power-law regression distribution in an ecological network signifies the existence of a multitude of uncommon species inside the network, thereby exerting a substantial impact on the structure and functioning of the ecosystem.

CHAPTER 3

RESULTS

3.1Network Inference

This study's network analysis and visualization component utilized a range of specialized R packages, including igraph, ggraph, tidygraph, and sna, to methodically evaluate the complex network of microbial interactions within the studied ecosystems. The first stage encompassed the calculation of fundamental metrics for network statistics. These metrics included degree centrality, which measures the number of connections each node has; closeness centrality, which determines how close a given node is to others; betweenness centrality, which identifies key nodes that enable network connectivity; and eigenvector centrality, which assesses a node's influence based on nearby nodes. Following this, a process of transforming species pairs, spice1 and spice2, was carried out, resulting in conversion into numerical labels that enabled standardized and effective network analysis. One crucial aspect of this phase is the smooth incorporation of each species' Ecological Importance Value Index (IVI) into the numerical framework. The Index of Ecological Significance (IVI), a measure used to assess the importance of ecological factors, was carefully linked to its respective network node, hence boosting the level of interpretation in subsequent visualization endeavors.

The visualization procedure was implemented using the ggraph package. The classification system utilized in this study is based on the concept of phyla, and it relies on a set of 21 unique colours to distinguish between various nodes. The colour gradient employed in this depiction effectively represents the distribution of various Phyla throughout the network, ranging from red to blue. Furthermore, the size of the nodes was carefully modified about their individual IVI scores, resulting in the creation of a visually meaningful gradient. In addition to aesthetic distinctiveness, the visualization was enhanced through the purposeful use of nuanced indicators, which provided the visualization with several layers of insights. The manipulation of node opacity, which reflects their calculated weighted scores in crucial centrality metrics such as degree, closeness, and betweenness, enabled the identification of nodes that had increased relevance. The size attribution of nodes played a crucial role in the visualization process, as it served as a visual indicator that conveyed supplementary information under the calculated IVI values. The larger nodes served as visual indicators indicating greater IVI scores, providing a quick understanding of the ecological significance of species inside the network. The positioning of nodes in the visualization was determined using phi values, computed to assess correlations among the nodes. The utilization of smaller phi values indicated increased dependency among nodes, providing an unbiased depiction of relationships. The primary focus of the investigation revolved around the careful identification and examination of particular clusters of nodes that possess a higher level of influence inside the network. The clusters, which exhibited increased relevance, underwent further research, greatly enhancing our understanding of complex microbial dynamics. The concept of dynamics.

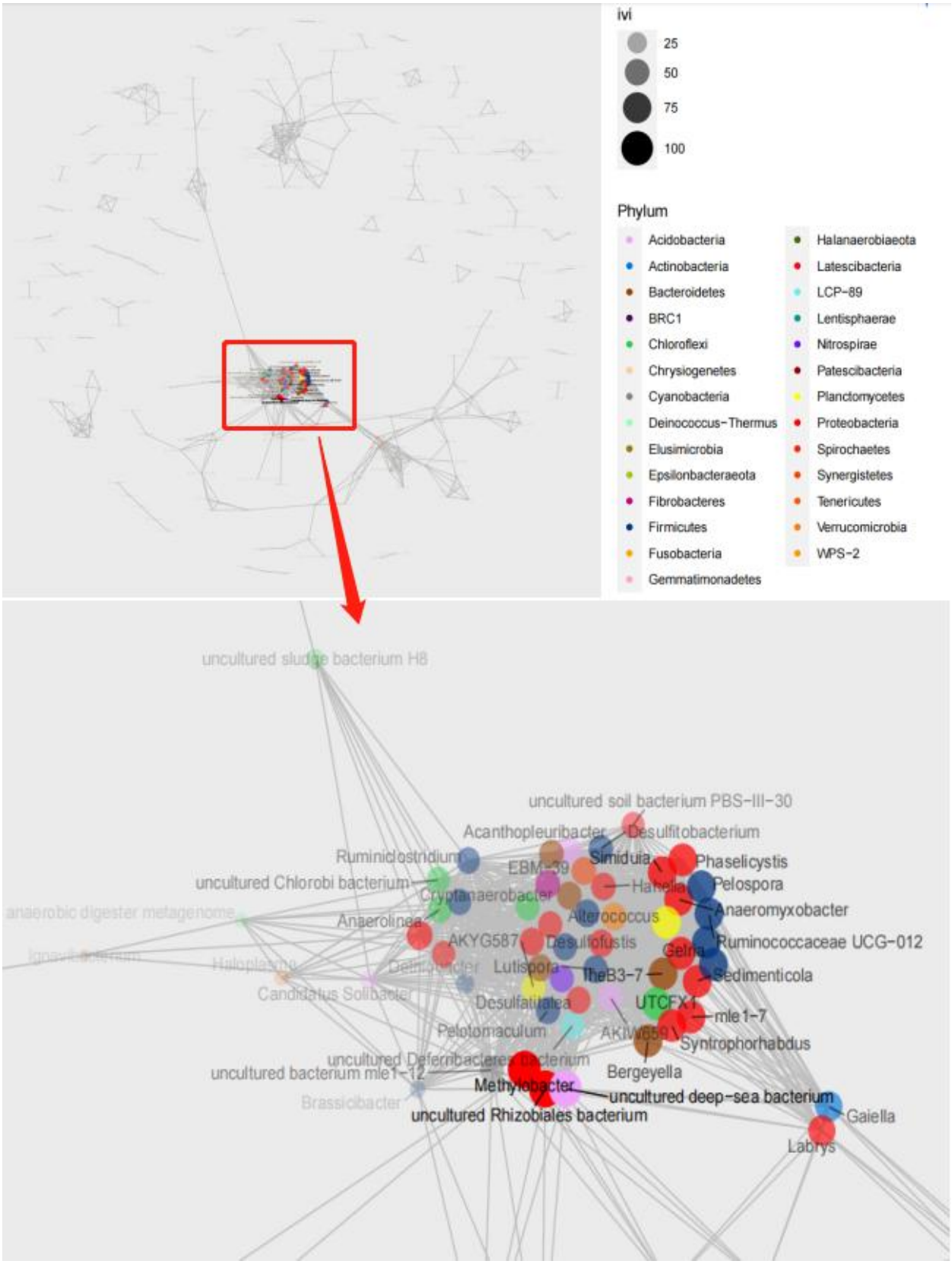


Figure 4:Microbiological network diagram of sample pit latrines in Vietnam

The figure demonstrates a notable prevalence of red and blue dots within the nodes. The red dots correspond to Proteobacteria, while the blue dots represent Firmicutes. This information is derived from the rating scale graph presented in Figure 7. Both microorganisms exhibit high scores on the Index of Vegetation Influence (IVI) and exert significant influence on the overall biome.

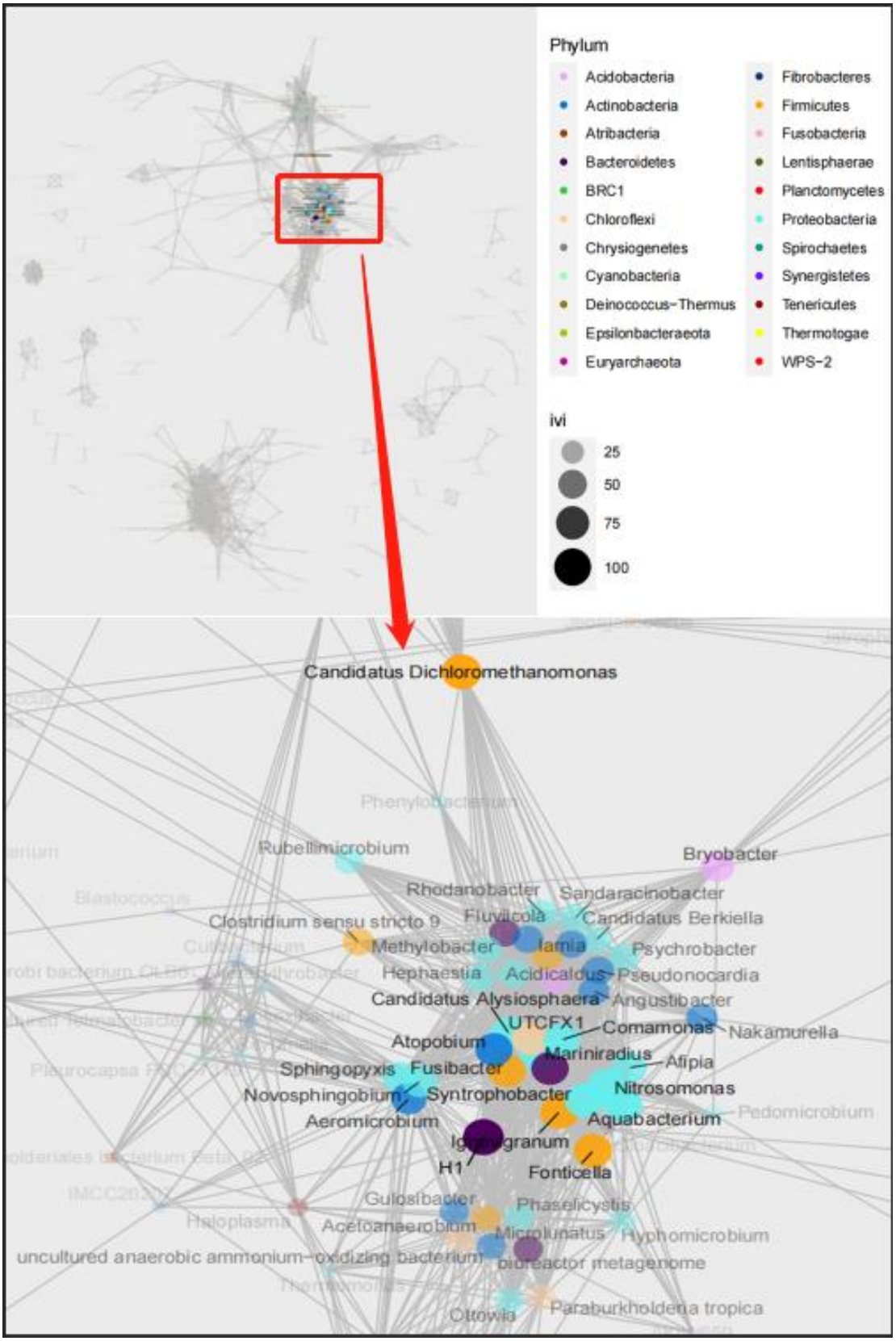


Figure 5:Microbiological network diagram of sample pit latrines in Tanzania

Among the network diagrams of Tanzania it can be noted that since the dominant Phylum is Firmicutes, the main diagram is mainly cyan in colour.

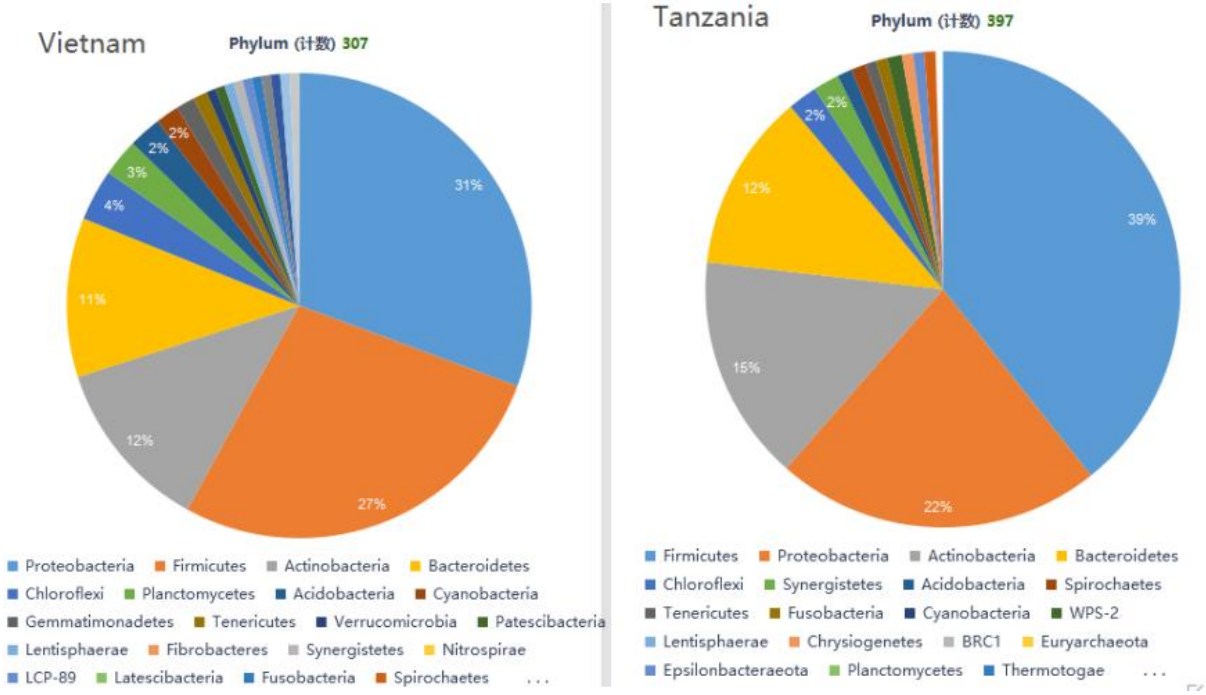


Figure 6:Pie chart of the distribution of the number of microbiological samples in the two regions

In the samples collected from the Vietnam region, 307 genera were identified, whereas in the samples obtained from Tanzania, this number was 397. The dominating Phyla in both regions were consistent, with Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Chloroflexi being the top 5 identified. In the context of Vietnam, it was observed that Proteobacteria constituted the most significant proportion, representing 31% of the samples. Conversely, in Tanzania, the proportion of Proteobacteria was comparatively lower, accounting for just 22%. In the samples collected from Vietnam, Firmicutes accounted for 27% of the total, making it the second most prevalent phylum. Conversely, Firmicutes exhibited the most significant proportion in Tanzania, comprising 39% of the samples.

In both regions, Actinobacteria, Bacteroidetes, and Chloroflexi were placed third, fourth, and fifth, respectively. This observation aligns with our anticipated findings, as Proteobacteria, Firmicutes, and Actinobacteria have been identified as prevalent taxa in pit latrines.

spreading_score	ivi	hubness_score	Phylum	Genus
100	100	62.28686076	Proteobacteria	Methylobacter
81.51995338	98.097413	75.10439828	Proteobacteria	uncultured Rhizobiales bacterium
81.51995338	98.097413	75.10439828	Acidobacteria	uncultured deep-sea bacterium
52.66005038	69.713658	82.82186284	Proteobacteria	Sedimenticola
52.66005038	69.713658	82.82186284	Proteobacteria	Phaselicystis
52.66005038	69.713658	82.82186284	Proteobacteria	Anaeromyxobacter
52.66005038	69.713658	82.82186284	Proteobacteria	mle1-7
52.66005038	69.713658	82.82186284	Planctomycetes	uncultured planctomycete
52.66005038	69.713658	82.82186284	Proteobacteria	bacterium enrichment culture clone Anammox_8
52.66005038	69.713658	82.82186284	Proteobacteria	Simiduia
52.66005038	69.713658	82.82186284	Bacteroidetes	IheB3-7
52.66005038	69.713658	82.82186284	Firmicutes	Ruminococcaceae UCG-012
52.66005038	69.713658	82.82186284	Chloroflexi	UTCFX1
52.66005038	69.713658	82.82186284	Firmicutes	uncultured Thermoanaerobacteraceae bacterium
52.66005038	69.713658	82.82186284	Bacteroidetes	Bergeyella
52.66005038	69.713658	82.82186284	Proteobacteria	Syntrophorhabdus
52.66005038	69.713658	82.82186284	Firmicutes	Pelospora
52.66005038	69.713658	82.82186284	Firmicutes	Gelria
86.33329251	59.503057	42.70056281	Actinobacteria	Gaiella
86.33329251	59.503057	42.70056281	Proteobacteria	Labrys

Figure 7:Top 20 microbiological IVI scores in Vietnam

Initially, a correlation may be discerned between IVI values and spreading scores. Within the dataset, the species Methylobacter, belonging to the phylum Proteobacteria, exhibits the highest IVI value of 100. This suggests that the Methylobacter gene is of utmost significance within the network, as evidenced by its high spreading and centrality scores.

Additionally, it is crucial to remark that several other genes have been uncovered within the Proteobacteria phylum. These genes include uncultured Rhizobiales bacterium, Sedimenticola, and Phaselicystis, demonstrating higher spreading and centrality scores. This suggests that the genes in the Proteobacteria phylum play vital roles in facilitating the dissemination and interconnectedness of the microbial network, potentially exerting substantial influence on interactions between microorganisms.

Furthermore, when considering their taxonomic categorization at the genus level, these genes' Inverse Simpson Index (IVI) values offer significant insights into their ecological significance. For example, even though the gene of the uncultured planctomycete in the Planctomycetes phylum demonstrates relatively high spreading and centrality scores, its lower IVI value suggests that it may have relatively little significance within the network.

Moreover, the dataset encompasses genes characterized by lower IVI values, including those belonging to the Actinobacteria phylum, such as the genera Gaiella and Labrys, as well as genes originating from other phyla, such as the Ruminococcaceae UCG-012 genus under the Firmicutes phylum. Although the IVI values of these genes are comparatively lower, they nonetheless exert significant functions within the microbial network.

spreading_score	ivi	hubness_score	Phylum	Genus
91.20968023	100	96.85300752	Bacteroidetes	H1
100	92.99606295	82.0112782	Firmicutes	Candidatus Dichloromethanomonas
87.20766703	86.1281787	87.14736842	Firmicutes	Ignavigranum
87.20766703	86.1281787	87.14736842	Proteobacteria	Aquabacterium
87.20766703	86.1281787	87.14736842	Proteobacteria	Nitrosomonas
87.20766703	86.1281787	87.14736842	Proteobacteria	Syntrophobacter
86.81774562	85.74322662	87.14736842	Firmicutes	Fonticella
70.08893353	79.28996296	100	Proteobacteria	uncultured Xanthobacteraceae bacterium
70.08893353	79.28996296	100	Firmicutes	Fusibacter
70.08893353	79.28996296	100	Proteobacteria	Candidatus Alysiosphaera
70.08893353	79.28996296	100	Bacteroidetes	Mariniradius
70.08893353	79.28996296	100	Chloroflexi	UTCFX1
70.08893353	79.28996296	100	Proteobacteria	Comamonas
70.08893353	79.28996296	100	Actinobacteria	Atopobium
71.76753169	70.2613732	86.36992481	Proteobacteria	Sphingopyxis
71.76753169	70.2613732	86.36992481	Actinobacteria	Aeromicrobium
71.76753169	70.2613732	86.36992481	Proteobacteria	Novosphingobium
79.76348691	69.2344807	76.45338346	Proteobacteria	Afipia
90.84990338	61.45274911	59.37857143	Acidobacteria	Bryobacter
71.26654833	50.95980691	62.74473684	Actinobacteria	Nakamurella

Figure 8:Top 20 microbiological IVI scores in Tanzania

In the Tanzanian region, the Genus "H1" from the Bacteroidetes phylum has the most effect, as indicated by its spreading_score and hubness_score of 91.20968023 and 96.85300752, respectively. Consequently, it achieves an overall IVI score of 100. Subsequently, the scores exhibit a progressive decline without a rapid decrease, as reported in the Vietnam region. The Genus holds significant influence and primarily falls within the phyla Firmicutes and Proteobacteria. Significantly, the Bacteroidetes phylum is home to Paludibacteraceae H1, often recognized as the most influential microbe.

Moreover, it is worth noting that genes belonging to the Aquabacterium, Nitrosomonas, and Syntrophobacter species, which fall under the Proteobacteria phylum, have elevated spreading and centrality scores. This implies that these genes have a pivotal function in the dissemination and interconnection of microorganisms within the microbial network, exerting substantial impacts on the interactions among microorganisms.

Concurrently, in conjunction with the taxonomic classification at the genus level of these genes, the IVI values offer valuable insights into their ecological significance. For example, while the Mariniradius gene, which belongs to the Planctomycetes phylum, has higher spreading and centrality scores, its lower IVI value suggests a relatively lower level of relevance within the network.

Furthermore, the dataset encompasses genes that exhibit lower IVI values, namely Atopobium and Nakamurella within the Actinobacteria phylum, along with Bryobacter under the Acidobacteria phylum. Although the IVI values of these genes are very modest, they still significantly impact the microbial network.

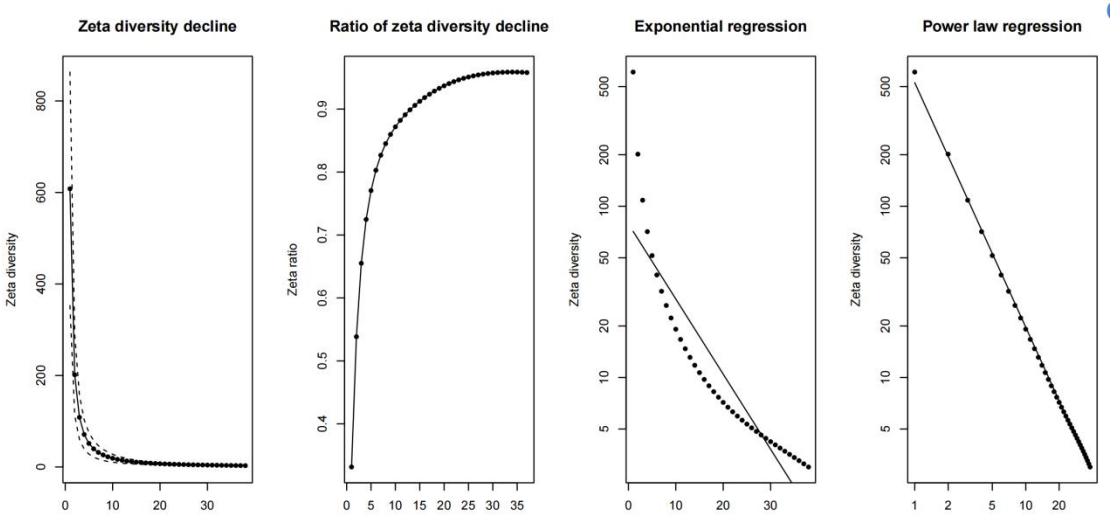


Figure 9:Distribution of microbial zeta diversity in pit toilets in Vietnam, Ratio of zeta diversity distribution and Regression Modeling in Vietnam

The Zeta diversity drop curve demonstrates a notable decrease throughout the early range of orders, specifically from 1 to 10. As the order of magnitude surpasses 20, the fall in Zeta diversity exhibits a progressive stabilization and tends towards infinity. This observation suggests a significant negative correlation between the order of samples and the abundance of coexisting microorganisms, with a notable reduction as the order increases. Nevertheless, as the order falls within the range of 10 to 20, there is a noticeable deceleration in the decline in Zeta diversity. This suggests a progressive decrease in the rate at which the number of shared species decreases. When the order surpasses 20, the decline in Zeta diversity remains generally stable and approaches zero, indicating that the quantity of shared species is significantly reduced beyond this threshold.

The Zeta diversity decline ratio demonstrates that the curve experiences rapid expansion throughout the range of orders from 1 to 10, followed by a gradual decrease in its slope. This observation indicates a notable decrease in the number of species shared within this stage. On the other hand, there is a noticeable decrease in the fall rate as the rank increases. The curve demonstrates a more consistent and gradual rise when the value is within the interval of 10 to 25 once the order exceeds a quantity of 25, the curve reaches a state of stability, suggesting that as the order increases, the disparities in species composition diminish gradually. The count of species shared between samples remains essentially constant.

The analysis of curve fitting reveals that the Zeta diversity drop curve of the samples demonstrates a more robust fit with the Power Law Regression model. When the number of orders exceeds 10, a significant proportion of samples aggregate towards the curve's tail, indicating a pronounced long-tail phenomenon. The elongated appendage depicted here signifies a taxonomic assemblage, including multiple species exhibiting diverse population sizes. This observation implies that as the order increases, there is a gradual decrease in Zeta diversity and a stable number of coexisting species. This finding further supports the notion that higher order is associated with a relatively favorable migration rate of species, which aligns with the conclusion derived from the Ratio of the Zeta diversity decline curve.

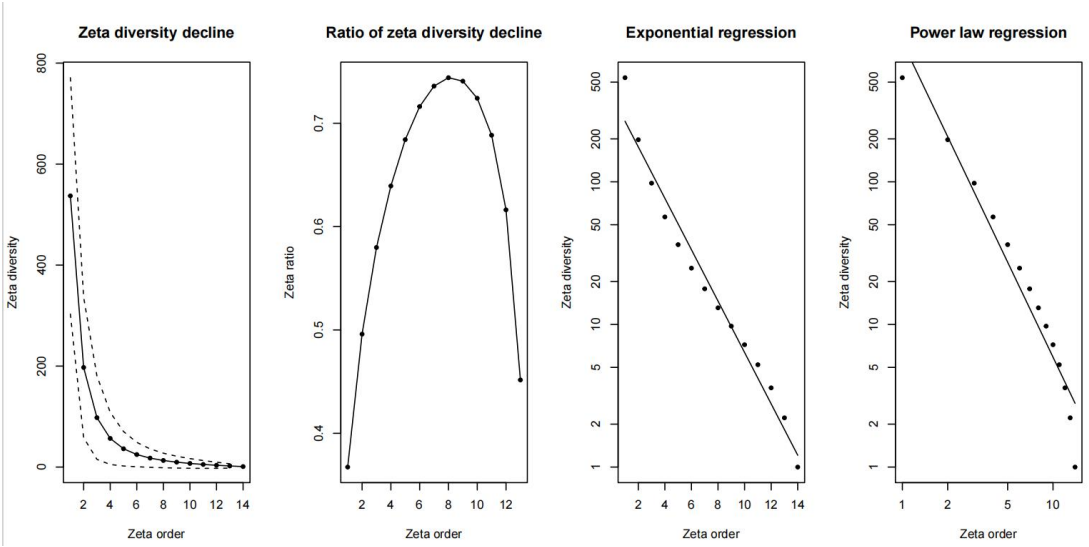


Figure 10:Distribution of microbial zeta diversity in pit toilets in Vietnam, Ratio of zeta diversity distribution and Regression Modeling in Tanzania

Zeta diversity is essential for comprehending species turnover, and the loss pace is shared species as taxonomic precision increases. The Zeta diversity values in the Tanzanian region demonstrate a notable decreasing trend. A significant and expeditious decline exists in the number of species shared across taxonomic orders 1 and 4. This observation implies a considerable species overlap across Tanzania samples at broader taxonomic levels. Nevertheless, when examining more detailed taxonomic classifications (namely, orders 5 to 12), the pace of Zeta diversity decline becomes less rapid.

Once the taxonomic order reaches level 12, the Zeta diversity values exhibit a trend of stabilization, suggesting that the quantity of species shared among taxa is relatively consistent at this threshold and higher.

The Vietnamese samples have a distinct pattern in Zeta diversity when compared. The Zeta values observed in Vietnam exhibit a pronounced and extensive decrease across a broader spectrum of taxonomic orders, ranging from 1 to 10. The data indicates that the samples collected from Vietnam exhibit a higher rate of species turnover, and there is a more noticeable decrease in the number of shared species as the data is examined at increasingly exact taxonomic levels.

Furthermore, significant variations between Tanzania and Vietnam can be observed in the zeta diversity decrease curves. In the Tanzanian setting, it can be noticed that there is a steady increase in the Zeta ratio values as one progresses through taxonomic orders 1 to 8. The observed rise in Zeta ratio values is simultaneous with a significant decrease in species overlap between these taxonomic groups. However, the scale of the decrease steadily decreases as we move up to higher taxonomic orders (orders 9 to 12). It is worth mentioning that when the taxonomic order surpasses 8, there is a significant decline in the ratio values. This decline indicates an increased loss rate in the diversity of species shared among different taxa.

On the other hand, within the specific framework of Vietnam, there is an observable upward trend in the decline of the zeta diversity ratio, eventually reaching a point of convergence at a

value of 1 as the taxonomic order is enhanced. This observation suggests a more even distribution of shared species across various taxonomic levels, indicating a reasonably stable species turnover.

Moreover, it is observed that the Zeta diversity distribution in the Tanzanian region exhibits little variations in the goodness of fit when comparing the Power Law Regression and Exponential regression models. This implies that the data from Tanzania does not exhibit discernible structural characteristics.

In conclusion, the research demonstrates that the Tanzanian region has a more pronounced decrease in Zeta diversity, particularly in the lower-order spectrum, compared to Vietnam. The contrasting trends observed in the zeta diversity decline curves underscore the considerable dissimilarity between the two regions. Additionally, it is worth noting that the data from Tanzania does not exhibit distinct structural characteristics since both the Power Law Regression and Exponential Regression models yield comparable fits to the Zeta diversity distribution.

CHAPTER 4

CONCLUSION AND DISCUSSION

Based on the findings presented in Chapter 3, it is evident that the samples collected from Vietnam exhibit a total of 307 Phyla. However, the samples from Tanzania display a count of 397 Phyla. A noteworthy disparity exists in the relative abundance of alkaliphilic bacteria seen in the samples obtained from Vietnam compared to those obtained from Tanzania. This implies that the introduction of lime powder into Vietnamese toilets may have eliminated a subset of acidophilic microorganisms that exhibit superior adaptation to acidic conditions.

Another notable disparity lies in the Firmicutes ratio observed between the two locations. In the context of Vietnam, the proportion of Firmicutes is 27%, whereas, in Tanzania, it is 39%. The observed discrepancy can be ascribed to the dietary practices of individuals in Tanzania who predominantly adhere to a vegetarian diet. Consequently, their fecal matter exhibits elevated quantities of dietary fiber in contrast to the populace in Vietnam. Firmicutes have been identified as beneficiaries of dietary fibre as a significant source of energy (Sun et al., 2022).

The increased abundance of Proteobacteria in Vietnamese toilets compared to Tanzania can be attributed to the substantial prevalence of animal faeces, particularly from cattle that consume plant-based feeds with relatively high protein content. Proteobacteria exhibit a pronounced capacity for protein decomposition, a process facilitated by the high protein content seen in animal faeces.

Paludibacteraceae H1 exhibits the greatest Inverse Simpson's Diversity Index (IVI) score, suggesting its substantial impact on the microbial community. The involvement of Paludibacteraceae H1 in the first hydrolysis and acid generation processes is crucial, as it facilitates the breakdown of proteins, carbohydrates, and fats into more basic molecules such as amino acids, sugars, and fatty acids. These simpler molecules serve as metabolic substrates for other bacteria (Satinover, 2020). Moreover, the protein within the metabolic substrates of Paludibacteraceae H1 gives rise to a competitive relationship with Proteobacteria, resulting in the inhibition of Proteobacteria populations in Tanzania, where Paludibacteraceae H1 has a predominant impact. However, it is observed that Paludibacteraceae H1 does not exhibit the same level of prevalence in Vietnam. This can be attributed to the prevalent use of lime as a disinfectant in Vietnamese toilets, which leads to a rise in pH levels. Consequently, the growth and proliferation of Paludibacteraceae H1, which thrives in acidic conditions, are inhibited.

In the context of Vietnam, Methylobacter has the most elevated Importance Value Index (IVI) score. Methylobacter exhibits a ubiquitous presence across diverse environmental contexts, including soil, aquatic ecosystems, and the surfaces of plants. The organism performs a vital ecological function through its symbiotic interactions with plant roots and leaves. Herbivorous organisms also consume plants that harbor Methylobacter, resulting in a notable occurrence of Methylobacter within their excrement. In Tanzanian culture, there is a prevalent adherence to a vegetarian dietary pattern. However, it is worth noting that the act of cooking and food processing commonly leads to a reduction in the presence of Methylobacter within the faecal

matter of individuals. Methylobacter depletes oxygen through its metabolic processes, creating a conducive anaerobic milieu for many anaerobic bacteria's proliferation and metabolic activities. In reciprocation, anaerobic microbes provide Methylobacter with methane, formic acid, formaldehyde, and more carbon sources, establishing a symbiotic connection that is mutually advantageous.
Based on the findings derived from the Zeta distribution, it can be inferred that the microbial distribution inside the samples adheres to a power-law distribution pattern. As the distribution order grows, there is a negligible impact on the distribution of microorganisms. This observation implies that the microbial ecosystem is experiencing substantial environmental stress. The microbial distribution seen in the samples collected from Tanzania, which conforms to a power-law distribution lacking substantial structural characteristics, can potentially be linked to the absence of external stimuli. The observed microbial distribution difference in the samples collected from Vietnam is plausibly attributed to the alteration in the acid-alkaline environment resulting from applying lime for disinfection purposes in the toilet facilities. The alteration, as mentioned above, exerts a significant amount of stress on the microorganisms, resulting in the emergence of a power-law distribution pattern.

The observed tendency also signifies the heightened mobility of organisms as order increases, facilitating the intermingling of species among various depths for microorganisms. Consequently, the alterations in the diversity of communal microbes exhibit negligible variation. On the other hand, in Tanzania, this particular attribute could be more readily observable, suggesting a lower level of microbial movement and limited capacity for species interchange across varying depths. In summary, examining the samples obtained from Vietnam and Tanzania reveals that the microbial distribution within these regions exhibits distinct patterns. The application of lime disinfection in Vietnamese toilet facilities has impacted the distribution of microorganisms, resulting in a power-law distribution pattern. In contrast, the absence of discernible structural characteristics in the microbial distribution observed in Tanzanian samples indicates a restricted microbial movement and a reduced capacity for species interchange, leading to a unique pattern compared to Vietnam.

Moreover, upon examining the samples obtained from Vietnam and Tanzania, distinct ecological disparities emerge, encompassing changes in microbial composition, dietary impacts, and environmental elements. These disparities lead to discrepancies in microbial communities across the two geographical regions. The prevalence of alkaliphilic bacteria in Vietnam suggests that lime has influenced the acidic environment, potentially leading to alterations in the structure of the microbial community.

Furthermore, it is plausible that the variations seen in the relative abundance of Firmicutes and Proteobacteria between the two locations can be attributed to dietary practices. Specifically, individuals in Tanzania predominantly adhere to vegetarian diets rich in dietary fibre, promoting the proliferation of Firmicutes. On the other hand, it is worth noting that in the context of Vietnam, the elevated protein levels seen in animal faeces have the potential to influence the prevalence and proliferation of Proteobacteria significantly.

In summary, the results underscore the significance of environmental conditions and nutritional influences on microbial populations across various geographical areas. The observed power-law distribution pattern in Vietnam implies a positive correlation between microbial movement and species exchange as the order increases. Conversely, the pattern observed in Tanzania shows

restricted microbial mobility and resistance to species exchange. The ecological differences between the two regions are responsible for the varied microbial compositions seen.

CHAPTER 5

FUTURE WORK

In order to enhance comprehension of the intricate microbial interactions and biodiversity within pit latrines, it is recommended that forthcoming research endeavors incorporate co-culture investigations. These investigations entail implementing regulated laboratory procedures in which various microorganisms are cultivated collectively within controlled environments. By examining the reciprocal effects of various microbial species on one another's development, metabolism, and behavior, researchers can learn about symbiotic connections, syntrophic interactions, and cross-feeding occurrences within microbial communities.

Co-culture experiments are valuable tools for gaining insight into the functions of individual microorganisms in the decomposition of pit latrine waste and their possible effects on the overall dynamics of microbial communities. Researchers can find meaningful interactions contributing to waste breakdown and nutrient cycling by cultivating different combinations of microorganisms obtained from pit latrine samples. This understanding can facilitate the formulation of specific approaches to improve waste breakdown and optimize the performance of pit latrines.

In addition, incorporating metagenomic and metatranscriptomic analysis into co-culture investigations can yield a thorough understanding of the functional capabilities and gene expression profiles shown by microbial communities residing in pit latrines. The investigations mentioned above can unveil the precise metabolic pathways and enzymes implicated in the decomposition of organic matter, hence providing insights into the fundamental mechanisms that govern microbial interactions and community dynamics.

Besides conducting co-culture experiments in laboratories, field investigations that involve on-site monitoring of pit latrines can offer valuable insights into the response of microbial populations to actual environmental circumstances and nutritional inputs. Conducting longitudinal sampling of pit latrines over time enables the documentation of seasonal fluctuations and enduring changes in the microbial community's composition and diversity. This methodology will facilitate the identification of temporal patterns in microbial interactions and the corresponding reactions to fluctuations in environmental variables.

In order to enhance co-culture and field studies, sophisticated network analysis methods can be utilised to deduce microbial interaction networks within pit latrines. These networks can unveil the complex interconnections among microbial species, encompassing mutualistic, competitive, and commensal interactions. By examining network features such as node centrality and modularity, scholars can discern keystone species that possess pivotal functions in upholding the stability and operational efficiency of the microbial community.

In summary, it is recommended that the following research endeavors prioritise investigations into co-culture studies, metagenomic analysis, network analysis, and field studies to comprehend microbial interactions and biodiversity within pit latrines thoroughly. The findings derived from these investigations will not just enhance our comprehension of microbial ecology in sanitation

systems but also provide a valuable contribution to establishing sustainable and effective pit latrine management strategies.

REFERENCR

Bartram, J., & Cairncross, S. (2010, November 9). *Hygiene, Sanitation, and Water: Forgotten Foundations of Health*. Hygiene, Sanitation, and Water: Forgotten Foundations of Hea lth | PLOS Medicine. https://doi.org/10.1371/journal.pmed.1000367

Graham, J. P., & Polizzotto, M. L. (2013, May). Pit Latrines and Their Impacts on Groundwater Quality: A Systematic Review. *Environmental Health Perspectives*, 121(5), 521–530. https://doi.org/10.1289/ehp.1206028

The pit latrine paradox in low-income settings: A sanitation technology of choice or a pollution hotspot? (2023, March 31). The Pit Latrine Paradox in Low-income Settings: A Sanitation Technology of Choice or a Pollution Hotspot? - ScienceDirect. https://doi.org/10.1016/j.scitotenv.2023.163179

Torondel, B., Ensink, J. H., Gundogdu, O., Ijaz, U. Z., Parkhill, J., Abdelahi, F., Nguyen, V. A., Sudgen, S., Gibson, W., Walker, A. W., & Quince, C. (2016, February 15). Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. *Microbial Biotechnology*, *9*(2), 209–223. https://doi.org/10.1111/1751-7915.12334

Bartram, J., & Cairncross, S. (2010, November 9). *Hygiene, Sanitation, and Water: Forgotten Foundations of Health*. Hygiene, Sanitation, and Water: Forgotten Foundations of Health | PLOS Medicine. https://doi.org/10.1371/journal.pmed.1000367

Analysis of pit latrine microbiota reveals depth-related variation in composition, and key parameters and taxa associated with latrine fill-up rate - PubMed. (2022, September 23). PubMed. https://doi.org/10.3389/fmicb.2022.960747

World health statistics 2019: monitoring health for the SDGs, sustainable development goals. (2019, May 21). World Health Statistics 2019: Monitoring Health for the SDGs, Sustainable Development Goals. https://www.who.int/publications-detail-redirect/9789241565707

Lindholm, E., & Larsson, L. (2022, November 14). *UTFORMNING AV SJÄLVMONITORERING FÖR ATT ÖKA FYSISK AKTIVITET: En systematisk litteraturöversikt*. UTFORMNING AV SJÄLVMONITORERING FÖR ATT ÖKA FYSISK AKTIVITET : En Systematisk Litteraturöversikt. https://urn.kb.se/resolve?urn=urn:nbn:se:mdh:diva-60520

Mara, D., Lane, J., Scott, B., & Trouba, D. (2010, November 16). *Sanitation and Health*. Sanitation and Health | PLOS Medicine. https://doi.org/10.1371/journal.pmed.1000363

Caldwell, E. L., & Parr, L. W. (1937, September 1). *Ground Water Pollution and the Bored Hole Latrine*. OUP Academic. https://doi.org/10.1093/infdis/61.2.148

Faecal Sludge Management | IWA Publishing. (n.d.). Faecal Sludge Management | IWA Publishing. https://www.iwapublishing.com/books/9781780404721/faecal-sludge-management

World Health Statistics 2014. (2014, May 15). World Health Statistics 2014. https://www.who.int/news/item/15-05-2014-world-health-statistics-2014

Wichelns, D., Drechsel, P., & Qadir, M. (2015, January 1). *Wastewater: Economic Asset in an Urbanizing World*. Wastewater: Economic Asset in an Urbanizing World | SpringerLink. https://doi.org/10.1007/978-94-017-9545-6_1

The Genome Sequence of the Halobacterium salinarum Type Strain Is Closely Related to That of Laboratory Strains NRC-1 and R1 - PubMed. (2019, July 11). PubMed. https://doi.org/10.1128/MRA.00429-19

Sabogal, R. I., Medlin, E., Aquino, G., & Gelting, R. J. (2014, March 1). *Sustainability of water, sanitation and hygiene interventions in Central America | Journal of Water, Sanitation and Hygiene for Development | IWA Publishing*. Sustainability of Water, Sanitation and Hygiene Interventions in Central America | Journal of Water, Sanitation and Hygiene for Development | IWA Publishing. https://doi.org/10.2166/washdev.2013.130

McGeoch, M. A., Latombe, G., Andrew, N. R., Nakagawa, S., Nipperess, D. A., Roigé, M., Marzinelli, E. M., Campbell, A. H., Vergés, A., Thomas, T., Steinberg, P. D., Selwood, K. E., Henriksen, M. V., & Hui, C. (2019, August 14). Measuring continuous compositional change using decline and decay in zeta diversity. *Ecology*, *100*(11). https://doi.org/10.1002/ecy.2832

Bennett, J. R., & Gilbert, B. (2015, December 23). Contrasting beta diversity among regions: how do classical and multivariate approaches compare? *Global Ecology and Biogeography*, *25*(3), 368–377. https://doi.org/10.1111/geb.12413

Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks - PubMed. (2020, June 22). PubMed. https://doi.org/10.1016/j.patter.2020.100052

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S., & Bähler, J. (2014, August 25). *Proportionality: a valid alternative to correlation for relative data*. bioRxiv. https://doi.org/10.1101/008417

Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns - PubMed. (2014, November 1). PubMed. https://doi.org/10.1086/678125

Sun, Y., Zhang, S., Nie, Q., He, H., Tan, H., Geng, F., Ji, H., Hu, J., & Nie, S. (2022, July 12). Gut firmicutes: Relationship with dietary fiber and role in host homeostasis. *Critical Reviews in Food Science and Nutrition*, 1–16. https://doi.org/10.1080/10408398.2022.2098249

Satinover, S. J. (2020, December 11). *Uncovering Process Challenges and Performance Differences for Complex Substrate Utilization in Microbial Electrolysis Cells*. TRACE: Tennessee Research and Creative Exchange. https://trace.tennessee.edu/utk_graddiss/5808

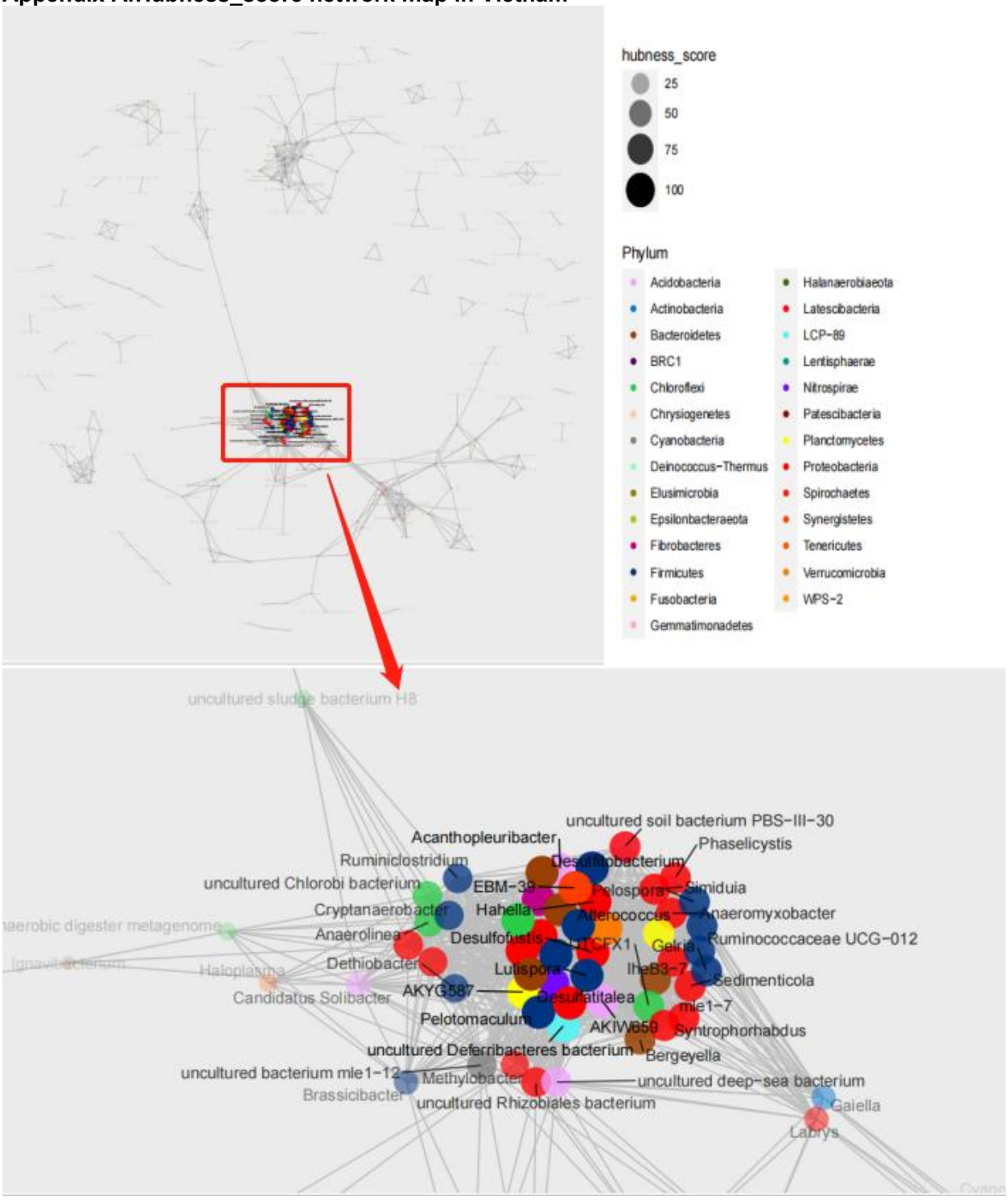
Zeta diversity - Wikipedia. (2018, October 1). Zeta Diversity - Wikipedia. https://en.wikipedia.org/wiki/Zeta_diversity

Soares Magalhães, R. J., Barnett, A. G. and Clements, A. C. A. (2011) "Geographical analysis of the role of water supply and sanitation in the risk of helminth infections of children in West Africa," *Proceedings of the National Academy of Sciences*, 108(50), pp. 20084–20089. doi: 10.1073/pnas.1106784108.

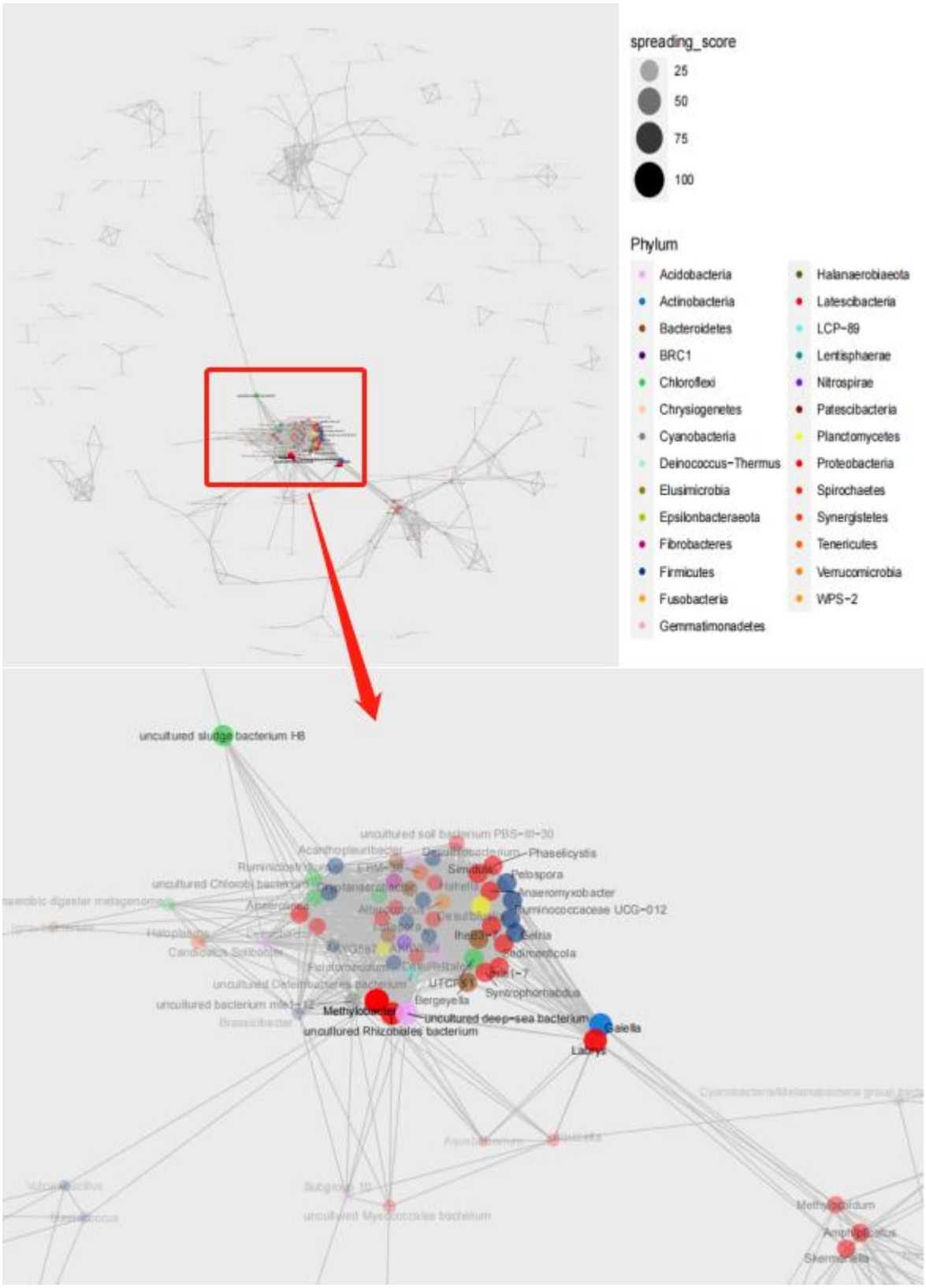
Rohwer, E. R. (2017). *The occurrence and wastewater chlorination fate of selected pharmaceuticals in South African surface water*. https://repository.up.ac.za/handle/2263/63363

APPENDIX

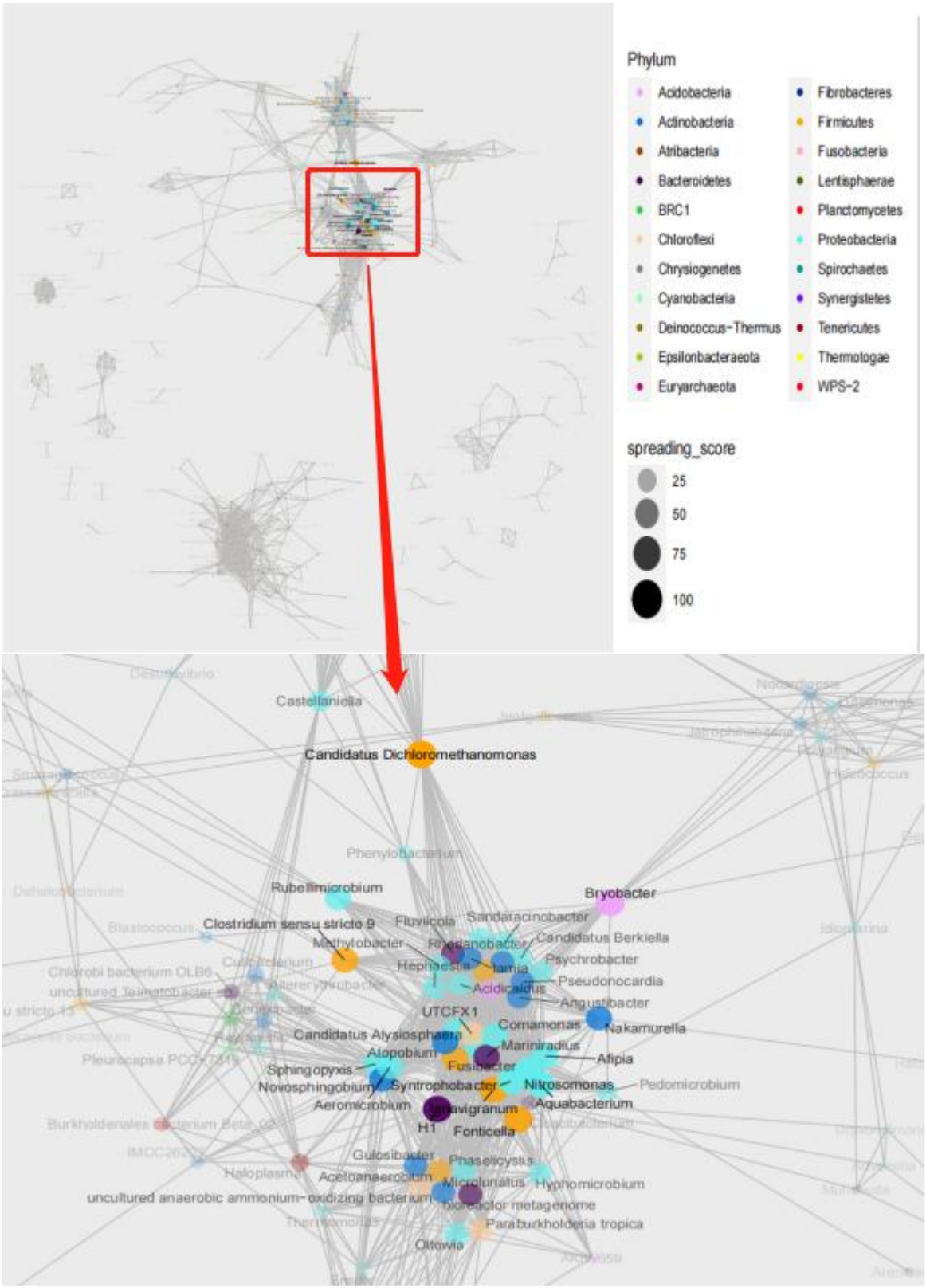
Appendix A:Hubness_score network map in Vietnam



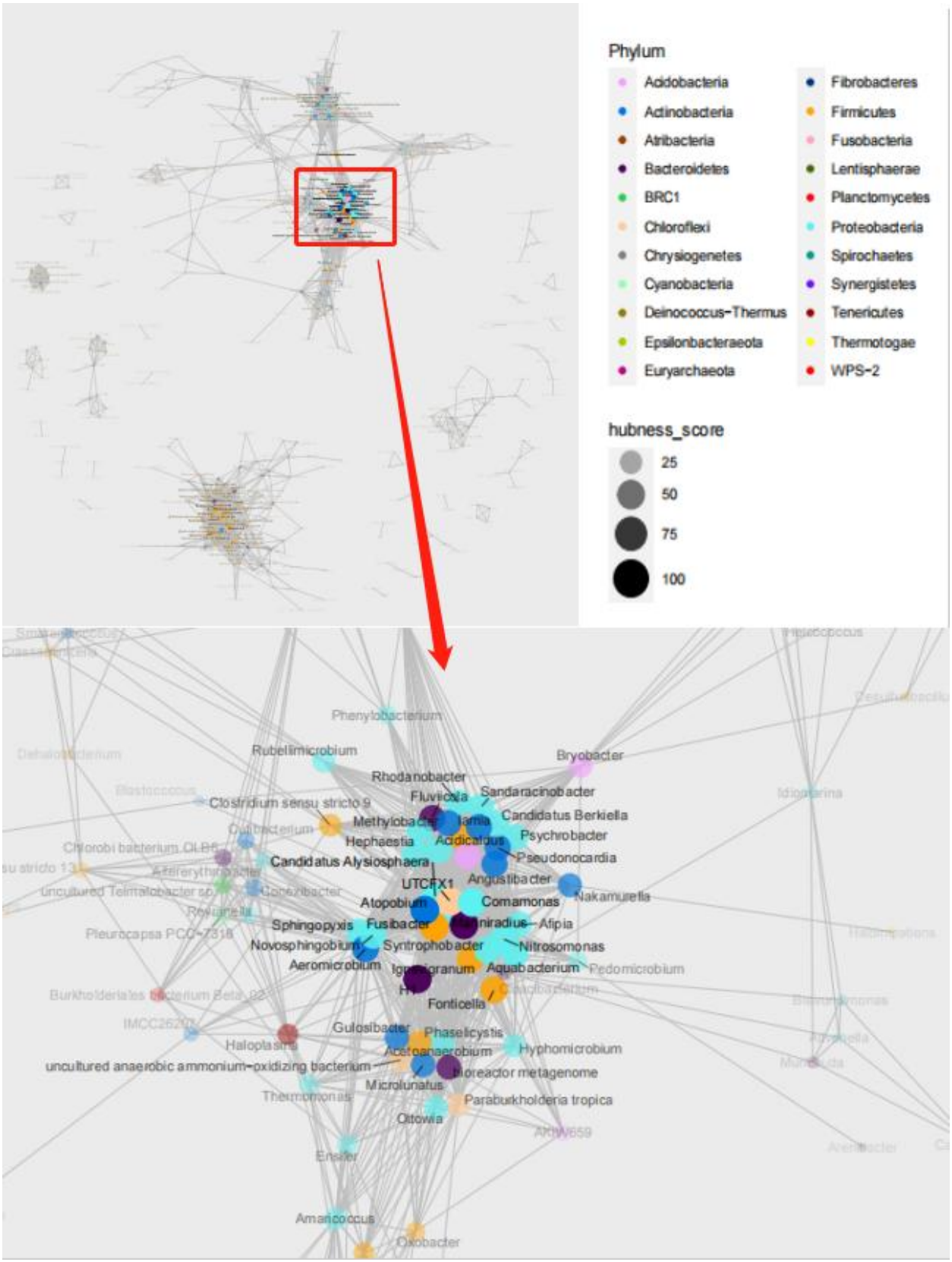
Appendix B:Spreading_score network map in Vietnam



Appendix C:Spreading_score network map in Tanzania



Appendix D:Hubness_score network map in Tanzania


Phylum
Acidobacteria
Actinobacteria
Atribacteria
Bacteroidetes
BRC1
Chloroflexi
Chrysiogenetes
Cyanobacteria
Deinococcus−Thermus
Epsilonbacteraeota
Euryarchaeota
Fibrobacteres
Firmicutes
Fusobacteria
Lentisphaerae
Planctomycetes
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Thermotogae
WPS−2
hubness_score
25
50
75
100
Phenylobacterium
Rubellimicrobium
Bryobacter
Rhodanobacter
Fluviicola
Sandaracinobacter
Clostridium sensu stricto 9
Methylobacter
Candidatus Berkiella
Hephaestia
Acidicaldus
Psychrobacter
Candidatus Alysiosphaera
Pseudonocardia
UTCFX1
Angustibacter
Atopobium
Comamonas
Nakamurella
Sphingopyxis
Fusibacter
Afipia
Novosphingobium
Syntrophobacter
Nitrosomonas
Aeromicrobium
Aquabacterium
Pedomicrobium
Fonticella
Gulosibacter
Phaselicystis
Haloplasma
Acetoanaerobium
Hyphomicrobium
uncultured anaerobic ammonium−oxidizing bacterium
bioreactor metagenome
Microlunatus
Thermomonas
Paraburkholderia tropica
Ottowia
Ensifer
Amaricoccus
Oxobacter

Appendix E:Comparison map of microbial Ratio of diversity between Tanzania and Vietnam