

**Community-based Analysis for Identifying Populations Relevant to Pollutant
Mitigation in Natural and Engineered Processes**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Lu Yang
August 2019

Copyright © 2015 by Lu Yang.

All rights reserved.

ACKNOWLEDGMENTS

First and foremost, I would like to express sincere thankfulness to my advisor, Dr. Qiang He. It is his encouragement along with my doctoral study that does this dissertation. He is the first person who leads my research path to microbiology study. I struggled for a long time in my first two years in my Ph.D. He always encourages the naïve me and gives enough patience and time in my study. He gives enough freedom in the research topic chosen and provides opportunities to the utmost of his capacity. It gradually increases my confidence and interests in research, and then I feel excited about studying the microbiome. When graduation is getting closer and closer, I gradually realize his past talking with me in weekly meetings. His logic thinking and academic rigor influence my thinking. However, it will take me more time to develop these abilities. For all of these, this gratefulness will be remembered.

I am lucky to have a great committee on my dissertation. It is really nice to have met Dr. Hilafu. It is lucky to have taken his class, which I think is very helpful for my data understanding. Besides, he is always supportive in my career and ready to help my research. Thanks to Dr. Schwartz. I value his opinions on my research and his support in the experimental sampling. I appreciate Dr. Zhuang, who arranges my transit from Nanjing University to UTK. Without his coordination, I will not be here. I also value his suggestions in my research.

My labmates are supportive of my research over the four years. Thanks to Dr. Si Chen, Dr. Kristen Wykoff, Dr. Yongfeng Wang, Liu Cao, Songyi Liu, Yabing Li, and Clifford Swanson. The friendship will be lasted forever. A special thanks to Dr. Si Chen, because of him, I have been able to encounter so many experimental data.

I also want to thank David Obenschain in Kuwahee wastewater treatment plant, who selfless helped my wastewater sampling.

I would like to express gratitude to my source funding from CSC DEEP100-Ph.D. program. Besides, I want to thanks to program colleagues, especially Xiaolong Liang, Fang Liu, Yongchao Xie, Chi Zhang, Qijun Zhang. In these four years, they are always ready to help me with life and research whenever I need.

I always feel gratitude to my master advisor in NJU, Dr. Jinhe Zhang. Wherever I am, he always cares about my research, career and my life.

Lastly, I would like to thank my parents, grandparents, and my boyfriend. Because of their endless love, I always believe I live like a happy girl. Thanks to my boyfriend's consideration and tolerance in these years from Nanjing to Knoxville. He always tries his best to fulfill my wish and to accompany me quietly. Also, his help in my fuzzy code debug is unparalleled.

ABSTRACT

Microorganisms are involved in various important environmental processes. While current understanding of these microbial processes is shaped to a large extent by studies of individual populations, increasing efforts have been made to understanding the roles of microbial communities as a whole in environmental processes, which is made possible with the development of high-throughput sequencing technologies. In this dissertation, the microbial communities in anaerobic waste treatment processes and stream waters influenced by anthropogenic activities were investigated as models of engineered and natural systems using microbial community-based analyses. In the anaerobic waste treatment processes, metagenomics analyses revealed the persistence of antibiotic resistant genes (ARGs) and association with specific microbial hosts, providing insight into potential targets for mitigating the spread of ARGs. Further, community-based analysis identified legacy effect as an important mechanism contributing to the assembly of microbial communities, shedding light on potential strategies for the control of important populations underlying waste treatment. In the investigation of stream waters impacted by anthropogenic activities, microbial community-based analyses enabled the successful identification of primary anthropogenic sources contributing to the microbial contamination in stream water, which has long been confounded using traditional indicator-based approaches. Results from this study provide an innovative technique for microbial source tracking not otherwise possible with individual population-based approaches. Community-based analyses, as demonstrated in this dissertation, are capable of identifying interactions between microbial populations which are essential for the survival, persistence, and function of microorganisms in the environment. Furthermore, community-based analyses are capable of utilizing all information embedded in microbial communities, which enables more precise and accurate quantification of microbial community composition and function, paving the way for the development of more effective data analytics techniques for the characterization and modeling of microbial communities.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1. Background	2
1.1 Microbial source tracking of pollutants	2
1.2 Pollutants mitigation by microbial populations	3
2. Dissertation Overview.....	4
Appendix A: Figures and Tables	5
CHAPTER 2 POLLUTANTS IDENTIFICATION BY METAGENOMIC APPROACH IN ENGINEERED PROCESS	6
Discovering Occurrence and Persistence of Antibiotic Resistance Genes in Anaerobic Digestion by a Metagenomic Approach.....	7
Abstract	7
1. Introduction	8
2. Materials and Methods.....	10
2.1 Experiment setup	10
2.2 DNA extraction and sequencing	10
2.3 Bioinformatics analysis.....	11
3. Results and Discussion.....	12
3.1 Occurrence, abundance, and diversity of ARGs	12
3.2 Taxonomy composition of the possible ARBs	13
3.3 Functional profile of the microbial communities.....	13
3.4 Co-occurrence of ARGs.....	14
3.5 Persistent of potential antibiotic-resistant bacteria	15
4. Conclusion	16
Appendix B: Figures and Tables.....	17
CHAPTER 3 POLLUTANTS MITIGATION IN ENGINEERED PROCESS	30
Section A: Legacy Effects Influence the Pathway Selection in Anaerobic Digestion	31
Abstract	31
1. Introduction	32
2. Materials and Methods.....	34
2.1 Dilution medium	34
2.2 Substrates and inoculum	34
2.3 Reactors set up	34
2.4 Sequencing processing of 16S rRNA gene.....	34
2.5 Data analysis	35
3. Results	36
3.1 Performance of the continuous reactors during formate spiking	36
3.2 Established <i>Methanosaeta</i> highly enriched in the formate spiked continuous reactors	36
3.3 Possible homoacetogens cooperated with established <i>Methanosaeta</i>	37
3.4 Legacy effects redirected the pathway selection	37
4. Discussion and Conclusion	39
Appendix C: Figures and Tables.....	41
Section B: Anaerobic Membrane Bioreactor Treatment for the Treatment of Recalcitrant Biofuel Process Wastewater.....	53
Abstract	53
1. Introduction	54
2. Materials and Methods.....	55
2.1 Wastewater pre-treatment	55
2.2 Anaerobic membrane bioreactor set-up	55
2.3 Chemical analysis	55
2.4 DNA extraction and 16S rRNA amplicon sequencing	55
2.5 Sequencing data analysis	56

3. Results	57
3.1 Test on compound composition for biofuel wastewater toxicity	57
3.2 Process adaptation to increased loading of toxicity biofuel process wastewater	57
3.3 Microbial community adaptation in anaerobic digestion	58
3.4 Microbial populations with a significant response to increases of wastewater loading	58
3.5 Methanogens involved in the anaerobic conversion of biofuel process wastewater	59
3.6 Syntrophic populations potentially involved in the anaerobic conversion of biofuel process wastewater	60
4. Conclusion	61
Appendix D: Figures and Tables	62
CHAPTER 4 POLLUTANTS SOURCE TRACKING IN NATURAL PROCESS	75
Section A: Rapid Change of Microbiome in Raw Sewage	76
Abstract	76
1. Introduction	77
2. Methods	78
2.1 Sampling sites selection and sample collection	78
2.2 DNA extraction and 16S rRNA sequencing	78
2.3 Dataset construction and statistical analysis	79
3. Results	80
3.1 Microbial community composition in sewage and human fecal samples	80
3.2 Effects of transportation and aging on sewage microbiome	80
3.3 Analysis of network characteristics and community/process variables	82
4. Conclusion	83
Appendix E: Figures and Tables	84
Section B: Stream Microbial Population in Response to Human Impact: A Perspective of Microbial Ecology	100
Abstract	100
1. Introduction	101
2. Methods	102
2.1 Sites description and sample collection	102
2.2 DNA extraction and DNA sequencing	102
2.3 Bioinformatics analysis and statistical analysis	103
3. Results	104
3.1 Bacterial community structure among samples	104
3.2 Human stress induces microbial AKP effects in streams	104
3.3 Unhappiness in AKP effects cause decreased microbial diversity	105
3.4 Potential influence of land use cause the AKP effects	106
4. Discussion and Conclusion	107
Appendix F: Figures and Tables	109
CHAPTER 5 CONCLUSIONS	125
LIST OF REFERENCES	128
VITA	143

LIST OF TABLES

Table 2-1 Summary of sample contribution to the assembly.....	18
Table 3-1 Microbial community clustered into cluster 1, cluster 2 and cluster 3.....	49
Table 4-1 Components in the 1L feed.....	62
Table 4-2 Organic compound concentration in AnMBR feed.	63
Table 4-3 Microbial community diversity during an episode of wastewater loading.	64
Table 5-1 Topological features of the empirical molecular ecological networks in comparison to random network.....	84
Table 5-2 Network hub, module hubs and connectors of microbial association networks.	85
Table 6-1 Results of MRM analysis to disclose the microbial community-related factors.	109
Table 6-2 Different levels of the developed area along the studying streams.	110

LIST OF FIGURES

Figure 1-1 Research workflow of water-energy-pollution nexus in this study.....	5
Figure 2-1 Analysis workflow for ARGs identification on metagenomic data.	17
Figure 2-2 Distributions of ARG types detected in digesters.	19
Figure 2-3 Venn diagram for the ARGs detected among the two treatments.....	20
Figure 2-4 The abundance of the ARGs subtypes in digester samples.....	21
Figure 2-5 Relative consistent microbial community among the ARG carried contigs at the phylum level.	22
Figure 2-6 Functional class of both samples in digester C and D.....	23
Figure 2-7 Functional class of defense mechanisms in digester C and D.....	24
Figure 2-8 Co-occurrence of ARGs and potential taxonomical origin (ARBs).	25
Figure 2-9 Summary of MGE abundances found in the metagenome samples.....	26
Figure 3-1 Figures of performance parameters.....	41
Figure 3-2 Barplot of methanogen relative abundance profile during the formate spiking.....	42
Figure 3-3 The relationships between full-length archaeal 16S rRNA gene sequences and different <i>Methanosaeta</i> in the digesters explained by neighbor-joining phylogenetic tree.....	43
Figure 3-4 Phylogenetic tree describes the bacterial composition in the digesters.	44
Figure 3-5 <i>Methanosaeta</i> correlated bacteria found during the formate treatment in the continuous digesters.....	45
Figure 3-6 Result of k-means clustering analysis of the taxonomic population.	47
Figure 3-7 The changing trend of the possible homoacetogenic bacteria with <i>Methanosaeta</i>	50
Figure 3-8 Phylogenetic distribution of the microbial community in newly formate batch reactors.	51
Figure 3-9 Methanogen composition comparison between acetate and formate treatment.....	52
Figure 4-1 Anaerobic membrane bioreactor (AnMBR) system design.	65
Figure 4-2 Pre-test of the biofuel wastewater.	66
Figure 4-3 Variation of COD removal rate and wastewater loading rate (% , vol/vol) in anaerobic digestion. .	67
Figure 4-4 Principal coordinates analyses of the samples.	68
Figure 4-5 Microbial community at phylum level in anaerobic digestion.....	69
Figure 4-6 Abundant bacteria with relative abundance greater than 0.5% found in any sample during anaerobic treatment.....	70
Figure 4-7 Microbial community change in anaerobic digestion.	71
Figure 4-8 Neighbor-joining phylogenetic tree showing the relationships of 16S rRNA gene sequences of <i>Methanobacterium</i> , <i>Methanosaeta</i> , and <i>Methanomassiliicoccus</i>	73
Figure 4-9 Possible utilization of influent organic compounds by syntrophic bacteria.....	74
Figure 5-1 Sampling points selected in the studied wastewater treatment plants and wastewater pumping stations.	86
Figure 5-2 Sample relatedness of wastewater and human fecal microbiome at the genus level.	87
Figure 5-3 Taxonomy distribution of top abundant families in the human fecal samples and the pooled sewage samples.....	88
Figure 5-4 Taxonomy mapping ratio of all classified families in the human fecal samples and the pooled wastewater samples.....	89
Figure 5-5 Alpha diversity measure of different types of wastewater samples.	90
Figure 5-6 Relationship between BOD ₅ and alpha diversity, VSS in wastewater samples.....	91
Figure 5-7 Comparison of pump station wastewater samples, influent wastewater samples and composite wastewater samples at the family level.....	92
Figure 5-8 Heatmap is showing the pattern difference among wastewater samples.	93
Figure 5-9 Biomarker detection analysis (LEfSe) on three types of sewage.....	94
Figure 5-10 Results of top abundant bacteria found as biomarkers in each type of wastewater by Lefse biomarker discovery analysis.	95
Figure 5-11 Subnetworks of each type of sewage.	96
Figure 5-12 Relation between network characteristics and microbial composition, environmental variables...	99
Figure 6-1 Study sites and the land cover of the developed area.....	111
Figure 6-2 Microbial population of streams and possible pollution sources at phylum level.	112
Figure 6-3 Bubble plots depicts the relative abundances of the main fecal sequences.	113

Figure 6-4 Sample relatedness of source and sink microbiome.....	114
Figure 6-5 Source contribution to urban streams estimated by SourceTracker.	115
Figure 6-6 Principle component analysis of samples relatedness, environmental variables and source contributions.....	116
Figure 6-7 Correlation between sewage contribution and coliforms in baseline.....	117
Figure 6-8 Correlation between sewage contribution and coliforms in baseline.....	118
Figure 6-9 Coliform concentration in waterbody and stormwater runoff.....	119
Figure 6-10 Alpha diversity of the microbial community in waterbody and stormwater runoff.....	120
Figure 6-11 Sample relatedness of stormwater samples and stream water samples.....	121
Figure 6-12 Log2Fold changes of taxonomic sequence variants between urban streams and reference streams.	122
Figure 6-13 Microbial community of streams and stormwater runoff at the family level.....	123
Figure 6-14 Site comparison of sewage inputs and nutrients concentration in urban streams.	124

CHAPTER 1

INTRODUCTION

1. Background

The new UN 2030 Agenda for Sustainable Development addressed the mission of reducing the negative impacts of urban activities. Urban activities produce wastes, in organic or inorganic format, are hazardous for human health and the environment. The reducing strategies focused on environmentally sound management, consisted the waste reduction and more efficient use of water and energy. Challenges in environmental management show the requirement of research on water-energy-pollution nexus (Figure 1-1). The nexus consists the synergies and trade-offs between water, energy, and pollution, which is still poorly understood (Kumar and Saroj 2014).

Water-energy is one of the interaction processes. Human activities, including industrial production, intensive farming, and daily human activities, require water and produce a large amount of wastewater. Industrial wastewater, livestock wastewater, and municipal wastewater are three important components of wastewater. The degradation and bioremediation of wastewater are still challenging since no reliable technique to date is capable of removing all these pollutants. However, anaerobic digestion together with functional microbial populations provides new directions in degrading these pollutants, such as toxic industrial wastewater, livestock wastewater with high organic loadings, and municipal wastewater. The anaerobic process can convert organic components in wastewater to biogas and help mitigate pollutants. The produced biogas can be directly applied in industrial, agricultural, and municipal activities. The biogas can also be transformed into electricity and then applied to human activities. Therefore, it is of great importance to research on the anaerobic conversion of organic matters to biogas in pollutants mitigation, an indispensable part in the water-energy process.

Water-pollution is another interaction process. Wastewater, together with stormwater, involve in the natural process, enters into the natural water. For example, during the transport of wastewater from households to wastewater treatment plants, wastewater pipeline leakage would increase the contamination of surface water. Another major cause in the deterioration of surface water quality is stormwater runoff, which is the pool of street solids and/or sewer-deposit contents. Both wastewater and stormwater contain a high concentration of fecal contamination, such as antibiotics, pathogens. The release of these fecal contaminations into surface water will cause various diseases in human beings and animals. It will also break the ecological balance of the surface water system. To protect public health and to maintain the surface water's ecological balance, much effort should be intensified in wastewater and stormwater management. Therefore, implement pollution source tracking is becoming vital in the water-pollution process.

1.1 Microbial source tracking of pollutants

Successfully pollution source identification is the premise of pollutants treatment, which boosts pollution mitigation more efficiently. Currently, the application of microbial indicators in detecting microbial pollution has been widely used. Most microbial pollution detection methods are pollutants-dependent. Currently, most widely concerned pollutants are antibiotics pollutants and fecal pollutants.

Antibiotics pollutants have been considered as a severe threat to public health. These antibiotics are originally used in human/animal disease treatment. The active compounds from the human/animal body will get excreted and will discharge to wastewater treatment plants. Therefore, wastewater treatment plants become the repository of antibiotics. Antibiotics can kill illness-causing bacteria and protect human/animals from infection. Antibiotic resistance is produced when bacteria develop the ability to survive in antibiotics treatment. These antibiotic-resistant developed bacteria are free to grow, diffuse, and make antibiotics not effective in disease treatment. Such resistance ability is caused by the acquisition of antibiotic resistance genes from other bacteria. For example, the pathogenic bacteria carry resistant genes and transport into the environment, such as soil and waterbody. Moreover, the horizontal transfer can help disseminate ARGs, which increases the risk of ARGs to the environment. The expansion of the resistance gene means the expansion of resistant bacteria type. It poses a risk to public health and increases the difficulty in disease treatment since some antibiotics cannot successfully treat specific disease anymore. Therefore, it is urgent for the scientists to find the antibiotic-resistant gene, antibiotic-resistant bacteria, understand the mechanisms that prevent the development and spreading of antibiotic-resistant bacteria. Then new antibiotics can be developed to treat the bacteria that cause disease and have resistant

potential. To date, most research have focused on revealing ARGs' prevalence during wastewater treatment, surface water (Martinez 2008), soil (Gothwal and Shashidhar 2015), lakes (Thevenon et al. 2012) of different regions of worlds. Due to the high prevalence and risk of ARGs existence, ARG elimination (Sharma et al. 2016), especially in wastewater treatment, has been highly paid attention to. The treatment efficiency related to degradable ARG has been discussed (Chen et al. 2013, Zhang et al. 2015b). However, little has focused on nondegradable ARGs. More studies on nondegradable ones during treatment will assist in risk evaluation associated with ARG to public health and ecosystems.

Shotgun metagenomics method has been regarded as a comprehensive tool for understanding the types and function of ARGs in environmental samples. This method sequenced the DNA from environmental samples and mining the antibiotic-resistant genes from the high coverage sequence data. Different from the most widely used qPCR method, it does not require specific primers designed from known target sequences. It is capable of detecting a more diverse ARGs, even the unknown ARG, and their functions. In addition, a variety of sophisticated analysis pipelines and redundant-removed databases have provided the convenience of downstream analysis of the sequences. Some databases, such as ResFinder (Zankari et al. 2012), CARD (Jia et al. 2017), ARDB (Liu and Pop 2009), etc., have been fully developed. Some algorithms, such as hidden Markov models(HMMs) (Yin et al. 2018), deep learning methods, machine learning methods, have been applied in the ARG predictions. However, most efforts are limited to the application of the qPCR method, or metagenomic shot reads blast analysis to date. Little has emphasized on assembled metagenomic reads, which is more accurate in ARG identification.

Fecal pollution is another serious concern. Even though the application of ARGs can successfully detect the antibiotic-resistant pollutants, but it is not able to differentiate pollution sources. Another method, host-specific PCR, is suggested to be capable of fecal pollution source identification. Specific PCR assay of some bacteria, especially the anaerobic bacteria, such as *Bacteroidetes* and *Firmicutes* that functionally dominate warm-blooded mammals' intestinal tract (Roslev and Bukh 2011), have been successfully applied in the fecal source tracking. The diversity of PCR assays involved in humans, bovine, cow, cattle, ruminant, swine, gulls, and others (Roslev and Bukh 2011). Its high repeatability and reproducibility make it more applicable (Ebentier et al. 2013). However, the recovery efficiency in the experiment process and the impact of PCR inhibitors still are the problems (Stoeckel et al. 2009). Also, no single genetic marker can be 100% precise specific and sensitive. With these limitations of qPCR, next-generation sequencing provides more opportunities in pollution source identification. A group of researchers has successfully applied this method, microbial community-based method, in microbial source tracking in environmental studies. The environmental objects including a variety of environmental waterbodies, such as oceans (Henry et al. 2016), lakes (Brown et al. 2017, Nakatsu et al. 2019), rivers (Wang et al. 2016). However, most of them focused on the source estimation for a specific site or emphasized the method development and verification (Ahmed et al. 2015, Staley et al. 2018). Overall general estimation overlooked the basic of source library definition. Moreover, most studies were case studies, and few of them have linked to the ecological aspects.

1.2 Pollutants mitigation by microbial populations

Anaerobic digestion, an engineered biological process, has been proposed as an environmental way to mitigate pollutants, such as animal wastewater and industrial wastewater. The wastewater contains a high concentration of organic compounds, which can be transferred into methane by a diverse of the microbial population. These microbial populations have been proved to play different functional roles in the four typical processes, hydrolysis, acidogenesis, acetogenesis, and methanogenesis. Among them, the most important procedure is methanogenesis since its slow biochemical reaction. To date, much research has discussed microbial functions in methane production. For examples, in mesophilic anaerobic condition, acetate-utilize methanogens include *Methanosaeta* and *Methanosarcina*. Methanol-utilize methanogens include *Methanococcoides* and *Methanosarcina*, while the rest of methanogens are H_2/CO_2 or formate-dependent. Such a complicated relationship between methanogens and substrates makes the understanding of the core microbiome much more complicated. More efforts are needed in understanding the microbial population in response to the substrate, which will contribute to the improvement of pollutants-energy conversion efficiency.

2. Dissertation Overview

To fill the knowledge gap related to antibiotics pollution in wastewater, fecal pollution source from wastewater and stormwater, and animal wastewater treatment and industrial wastewater treatment discussed above, we carried out the following three research tasks.

Task 1: Chapter 2 aims at studying the persistence of pollution biomarkers, antibiotic-resistant genes, and their potential hosts in anaerobic digestion by shotgun sequencing. Anaerobic treatment of animal wastewater was proposed as the engineered process. Comprehensive understanding of antibiotic-resistant bacteria and mobile genetic elements were also investigated.

Task 2: Chapter 3 aims at finding the core microbial populations that help improve the efficiency of bioconversion of pollutants to methane. In the first part of this chapter, the hypothesis was that formate, a crucial electron carrier in the final process of methane conversion, would be favorable to hydrogenotrophic methanogens under long term anaerobic treatment. 16S rRNA gene amplicon sequences compared the relevant microbial populations under formate treatment and recovered in the long-term anaerobic digesters. In the second part of this chapter, we engineered anaerobic membrane bioreactor to mitigate recalcitrant biofuel wastewater pollution. The microbial populations, capable of enriching microbial biomass but resistant to the toxicity, will be disclosed.

Task 3: Chapter 4 aims at tracking the pollution source in the natural environment by microbial community analysis. In the first part of this chapter, we designed to fill the knowledge gap in current microbial source tracking study, to clear the inner border of a vital urban pollution source, sewage, in source tracking. In the second part of this chapter, the hypothesis was that the microbiological changes induced by urban perturbations vary by pollution source, and therefore causes transitions of the stream from healthy to unhealthy states. Stream health condition was assessed by community-based microbial source tracking analysis. The microbial comparison between reference streams and urban streams were also implemented.

Appendix A: Figures and Tables

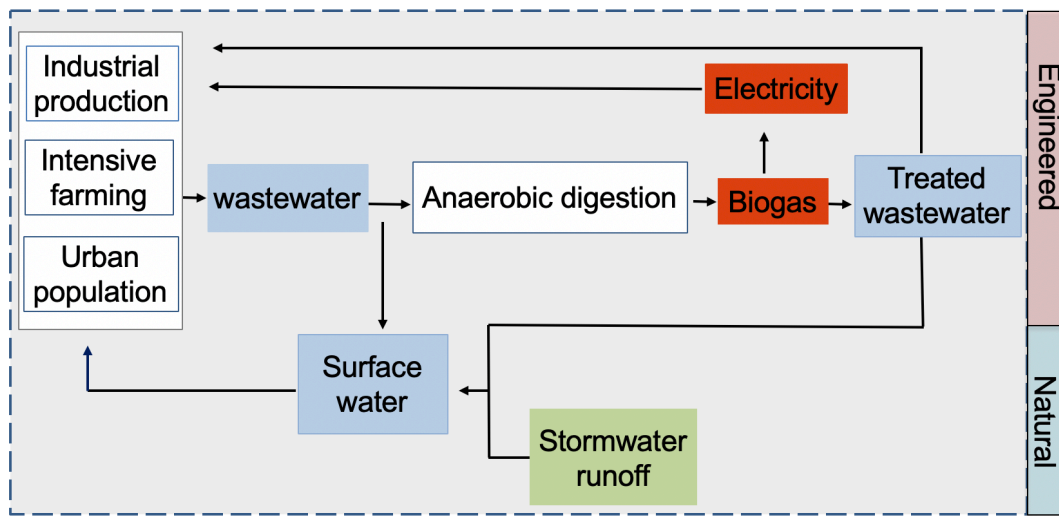


Figure 1-1 Research workflow of water-energy-pollution nexus in this study.

CHAPTER 2
POLLUTANTS IDENTIFICATION BY METAGENOMIC APPROACH IN
ENGINEERED PROCESS

Discovering Occurrence and Persistence of Antibiotic Resistance Genes in Anaerobic Digestion by a Metagenomic Approach

Abstract

Animal waste is one vital source of antibiotic-resistant genes (ARGs). Poor animal waste management can increase the burden of disease and environmental impact. Anaerobic digestion is an environmentally friendly method to treat organic wastes. Various studies have focused on degradable ARGs in anaerobic digestion, while few investigations have paid attention to non-degradable ARGs in anaerobic digestion. Therefore, this study aims at determining the persistence of ARGs and potential hosts of ARGs by shotgun metagenomic sequencing. Six anaerobic digestion samples, respectively from mono-digestion and co-digestion, were co-assembled into long contigs. High mapping ratio (> 68%) from the deep sequencing suggests the comprehensiveness of our investigation. ARG analysis discovered 126 ARG types, while the most abundant and persistent ones were *vanUG*, *dfrE*, *msrB*, *PmrE*, *LlmA*. Those ARGs types belonged to 20 ARG categories, mainly consisted of *macrolide-lincosamide-streptogramin*, *glycopeptide*, and *multidrug*. Moreover, the persistence of ARGs co-occurred with some bacteria mainly originated from *Firmicutes*, *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes*. The occurrence of potential ARBs was interestingly found persistently harbor multiple ARGs, and some ARGs could affiliate to several bacteria. Furthermore, we found some resistant functions persisted, which comprised of some transporters, resistant proteins, restriction enzymes, and multidrug efflux pump. Lastly, the persistence pattern was also characterized by persisted mobile genetic elements (MGEs).

Keywords: antibiotic resistant gene; metagenomics; persistence; mobile genetic function

1. Introduction

Antimicrobial agents have been used in veterinary and animal farming for the intensive development of farming, which increases the propagation probability of antibiotic resistant gene (ARG) in the environment. Veterinary antibiotics entering the environment in the form of using manure as fertilizer for land application is one of the main origins of antibiotics (Kim et al. 2012). The existence of ARGs causes a large number of concerns due to their potential environmental risk (Sun et al. 2018). It is alarming that these ARGs can be transferred into the environment by the processes of transduction (via bacteriophages), conjugation (via plasmid and conjugative transposons), and transformation (via incorporation into the chromosomal DNA plasmids and other DNA from other organisms) (Levy and Marshall 2004). After the horizontal transfer of ARGs, antibiotic compounds, antibiotic-resistant bacteria (ARB) and ARGs can colonize and proliferate in human, soil, or water and would threaten human and ecosystem health. Therefore, how to deal with these large quantities of ARGs-contained manure poses an urgent problem to waste management.

Anaerobic digestion, an environmentally sympathetic method in dealing with waste (Blake et al. 2017), provides the advantage of waste recycling and cleaner production. Especially anaerobic digestion has been identified to be capable of degrading ARGs. (Zhang et al. 2016) suggested that anaerobic co-digestion of food waste and sewage sludge could reduce the total abundance of ARGs, and the main driver for the fate of ARGs depends on the microbial community. The different mass ratio of swine manure with wheat straw could decrease the total abundance of ARGs (Song et al. 2017). Thermophilic anaerobic digestion with Fe^0 can help to degrade ARGs in the co-digestion of waste sludge and kitchen waste (Gao et al. 2017). However, most of these studies are focusing on what ARGs can be degraded during the anaerobic digestion, little has answered the question of what ARGs can persistently exist during the anaerobic digestion. Persistence of ARGs and the characteristic of stability (Barbosa and Levy 2000) means the capacity of populations survive under different survival conditions (Brauner et al. 2016). Persistence of ARGs needs to be gained much more attention since this characteristic hinders the process of ARG treatment. Therefore, to understand how ARGs persist in the anaerobic digestion can help improve the degradation of ARGs in biotechnology.

Metagenomic sequencing is a powerful technology in identifying a broad diversity of ARGs, which overcomes the drawbacks of the amplified based methods, such as quantitative polymerase chain reaction (qPCR) and amplicon sequencing. qPCR method has been widely used to identify the diversity and abundance of ARGs in former studies. A variety of ARGs were successfully found in different types of environment, such as treated and untreated wastewater (Mao et al. 2015b, Rodriguez-Mozaz et al. 2015, Xu et al. 2015), natural water (Czekalski et al. 2015), soil (Chen et al. 2016a), animal wastes (Czekalski et al. 2015, He et al. 2016). Especially, animal wastes, such as cattle manure and poultry waste, have been widely recognized as the reservoir of ARGs (Czekalski et al. 2015) with a high diversity of ARGs. However, the qPCR method can identify the specific ARG under the use of a specific primer, which is limited to the availability of market primers and the chosen of limited types of primers. Therefore, a desirable broader suite of ARGs could not be detected (Tien et al. 2017). Besides, 16S amplicon sequencing has been accompanied by the qPCR methods for identifying the ARBs. Possible hosts of ARGs were then determined by the pure statistical analysis under the correlation analysis. Moreover, low-throughput, amplification bias, statistically significant, false-positive, or false-negative results increase our thinking on the reliability of the found ARBs. And most current studies only identify the massive taxonomy composition without knowing the novel ARGs and their functions, since amplicon base methods are insufficient to provide the accurate function of the microbiome. Luckily, the metagenomic method can overcome above-mentioned drawbacks and can provide a more comprehensive profile of ARG diversity and its corresponding ARBs. Especially, a successfully metagenomic sequencing, annotation of short reads could help better survey the ARGs in the environment. A closer analysis by predicted open reading frames from assembled contigs to predict ARG could provide more accurate information on the ARGs identifications (Guo et al. 2017). Furthermore, mobility gene elements can be analyzed from the assembled contigs, which help further understanding the ARGs. Such a process has been mostly applied in the wastewater environmental system ARGs detection, while little has been studied for animal waste, especially during the anaerobic digestion.

Therefore, for surveying a broader diverse of ARGs, we used metagenomic sequencing approach, assembled the short reads into long contigs, and identified ARGs in the digesters in this study. In the meanwhile, ARGs potential hosts, ARBs, were also analyzed. We aimed to fully understand the occurrence, diversity and abundance of the ARGs and ARBs in the anaerobic digestion of different types of animal wastes. Besides, we aim at finding what ARGs and ARBs can persistently existed in anaerobic digestion. We hope this study could enhance our current understanding of the ARGs profile and the fate of ARGs in engineered treatment of animal wastewater.

2. Materials and Methods

2.1 Experiment setup

Triplicate mesophilic continuous anaerobic mono-digesters (digested with dairy manure, named C) and co-digesters (digested with dairy manure and chicken waste, named D) were respectively cultivated with inoculum from the former mentioned experiment (Wu et al. 2016b). Similar to previously described, the working volume of the digesters was 3.6 L and the operation temperature was 35 °C, the hydraulic retention time was 20 days. In the mono-digesters, the organic loading rate (OLR) was 1.0 g volatile solids (VS)/L/day. In contrast, the OLR was raised to 1.3 g VS/L/day in the co-digesters and gradually raised to 1.5 g VS/L/day by adding more poultry waste. During the operation process, the process performance of the digesters remained stable. Then we collected the biomass samples and stored them at -80 °C before DNA extraction.

2.2 DNA extraction and sequencing

Biomass and metagenomic DNA was extracted by grinding, freezing-thawing, SDS-based methods (Zhou et al. 1996). To summarize, for each time point, 6-ml of sludge sample for each reactor were collected. Biomass in the 6-ml sludge sample was pelleted in three Eppendorf tubes. Therefore, there are three tubes (with biomass) for each digester at each time point. The pellet from 0.5 mL of sludge typically yields 5ug of DNA (80-120 ng/μL in 50 μL). So, each tube of biomass in this shipment is expected to yield 20 μg DNA. Since each sample has three tubes of biomass, the total yield could reach around 60 μg for each sample. The metagenomic data was sequenced by Illumina HiSeq SE machines at Los Alamos National Laboratory.

For 16S rRNA sequencing, DNA extractions were performed using MP FastDNA™ Spin Kit for Soil according to the manufacturer's instruction (Protocol Revision #116560200-201608). Briefly, the biomass samples were resuspended in the 978 μl Sodium Phosphate Buffer solution and 122 μl MT Buffer, then followed by 100 seconds vortex on MP biomedical fast prep instrument. Then the DNA was purified with Zymo Genomic DNA Clean & Concentrator™-10 kit according to the instructions provided by the manufacturer. The DNA concentrations were determined using Thermo Scientific NanoDrop ND-3300 Fluorospectrometer. Polymerase chain reactions were prepared with a cocktail mix containing 12.5 μL Phusion flash Master Mix, 10 μL ultra-pure water, 1 μL forward primer, 1 μL reverse primer, and 2 μL (100 to 150 ng) DNA template. Each universal primer pairs were designed to target the V4 region of bacterial and archaeal 16S rRNA genes: forward primer 515F (GTGCCAGCMGCCGCGGTAA) with 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), and forward primer link (GT); reverse primer 806R (GGACTACHVGGGTWTCTAAT), with adapter (CAAGCAGAAGACGGCATACGAGAT), link (CC), and a unique 12-base specific barcode for each sequence. The PCR program included one cycle of 94 °C for 3 minutes, followed by 35 cycles at 94 °C for 45 seconds (denaturation), 55 °C for 1 minute (annealing), and 72 °C for 1 minute and a half (elongation), a final extension at 72 °C for 10 minutes, and then stored at 4 °C. After the program finished, the samples were run on an Agilent 2100 Bioanalyzer Instruments with Agilent DNA 7500 chips to ensure the amplicon quality and measure the DNA concentrations. Post PCR, amplicons were pooled (a set of 9 or 10) based on the peak height from Agilent DNA 7500 analyses and the concentrations were measured using KAPA Illumina kit for quantifying library (KK4824) from KAPA Biosystems. Samples were pooled again according to the results and number of sequences desired into a final one and the concentration was measured again with the KAPA quantifying kit. 4 nM pooled libraries were diluted with 10 mM Tris/0.05% tween buffer (pH = 8.5) and denatured with 0.2 N freshly made NaOH solution. The sample was then combined with 20% PhiX control kit to increase the library diversity. The mixture was incubated at 96 °C for 2 minutes using a heat block and then put into a water-ice bath for 5 minutes. After the sample was loaded, an Illumina MiSeq System with a MiSeq Reagent Kit v2 (300-cycles) and Metagenomic workflow was selected to execute the 16S protocol using the MiSeq Reporter software (MSR) at the University of Tennessee Genomics Core (UTGC).

2.3 Bioinformatics analysis

The bioinformatic strategy in the ARGs identification are described in the following steps, the workflow was shown in Figure 2-1 and the detailed summary of the assembly are shown in Table 2-1: (1) More than 44Gb paired-end reads for each sample, with 1 billion raw reads in 6 samples representing two digesters respectively in FASTQ format are quality filtered in *Trimmomatic v0.36* (Bolger et al. 2014) with sliding window size of 4, sliding window minimum quality of 15, minimum read length was set at 36, trailing minimum quality at 3, leading minimum quality at 3. As a result, 87% and 92% of the raw reads remained high-quality reads. (2) For achieving a better coverage, reads from 6 samples (duplicates in triplicates digesters) in type C and 6 samples in type D were respectively assembled. De novo co-assembly was performed respectively on two types of digester samples by *MEGAHIT* (Li et al. 2015), with a minimum k-mer varies from 1 to 127, maximum k-mer varies between 1 and 255, k-step between 1 and 28. For ensuring a reasonable length for open reading frames (ORF) prediction, all contigs with a length greater than 2000bp were remained, with 207,652 and 230,337 contigs in C and D respectively. For assembled contigs, N50 are 7,204 and 7,715 for C and D; N75 are 3,516 and 3,631 for C and D. Contigs length greater than 2000bp reach the total length of more than 1 billion for each digestion type. GC% contents of the assembly are 51.19% and 50.68% respectively in C and D digester. For the complexity of the anaerobic animal waste digestion samples, the percentage of reads that could be assembled into long contigs suggest a relative higher assemble rate than other research. (3) ORFs were predicted from contigs using *Prodigal* (Hyatt et al. 2010). (4) 16S rRNA genes from the contigs were predicted by *barnnap*. (5) ARGs were identified by *DeepARG* (Arango-Argoty et al. 2018b). The running parameters were set with the identity of 50, coverage of 50, e-value of 1e-10, the probability of 80%. Mobile Genetic Elements (MGEs) were filtered by a set of collected MGEs from National Center for Biotechnology Information (NCBI). The running parameters were set with the identity of 25, coverage of 50, e-value of 1e-10. (6) Taxonomy assignment of the ORFs was performed in *Kaiju* (Menzel et al. 2016). NCBI bacteria and archaea protein genomes were downloaded as the reference database. LCA (lowest common ancestors) algorithm was used in the taxonomy assignment. The running parameters were set with a minimum match length of 11 and a minimum match score of 65 under *mem* mode. (7) Cleaned raw reads were mapped to the assembled contigs in *BOWTIE* (Langmead et al. 2009) with end-to-end alignment type. Bam file produced in *BOWTIE* was sorted in *SAMtools* (Li et al. 2009), and the number of reads mapped to each contig was summarized in *SAMtools*. Each contig mapping rate was then used for the relative abundance of ARGs population normalization. (8) Protein sequences of the predicted genes were searched against the Non-Supervised Orthologous Groups (eggno-mapper, V4.5.1, mapping mode of DIAMOND, the rest settings are all default settings) (Huerta-Cepas et al. 2016) and the detected KEGG list was searched against Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Kanehisa and Goto 2000) by a self-written script in R.

For 16S rRNA Illumina Miseq sequencing data, pair-ended sequences were denoised under QIIME2 pipeline DADA2 algorithm. A mapping of feature sequences was summarized, and a pre-trained Naive Bayes classifier based on Greengene 13_8 99% OTUs was used in the taxonomy analysis (DeSantis et al. 2006).

3. Results and Discussion

3.1 Occurrence, abundance, and diversity of ARGs

Shotgun metagenomic sequencing was used to investigate the abundance of known ARGs by annotating the assembled contigs (> 2000bp) against the merged database from CARD, ARDB, UniProt, which contains 2203, 2128 and 10602 genes respectively. The output results consist of the antibiotic category to which a gene confers resistance and the antibiotic group to which the genes belong (Arango-Argoty et al. 2018a). Results indicated that a total of 694 assembled contigs (0.33%) of digester C and 890 assembled contigs (0.39%) of digester D were respectively found with 738 and 946 antibiotic resistant gene copies, assigned to 20 and 23 resistant categories respectively (Figure 2-2). Detected ARGs were all normalized by 16S gene copy numbers (ARGs/16S rRNA genes detected in digesters), which helped compare the ARG content from different samples (Yin et al. 2018). Mono-digestion had a total relative abundance of 5.55 ARGs/16S rRNA genes, while co-digestion had 6.10 ARGs/16S rRNA genes. It did not exhibit too many differences between the mono-digestion and co-digestion, which may suggest the addition co-digestion of chicken waste may not significantly improve the ARG abundance.

Antibiotic resistance genes can be grouped into the same antibiotic category may confer the same resistance profile and have the same action mechanism. As a result, 20 common resistance types, as well as multidrug resistance, were identified across mono digestion and co-digestion digesters (Figure 2-2), while the different three types expressed in acriflavine, streptothricin, and tunicamycin categories. Acriflavine previously has been found some bacteria in chicken feces can carry acriflavine ARG (Ma et al. 2016). Streptothricin has been regarded as treat the coliform bacteria of cecal contents in chicken (Moore et al. 1946). Tunicamycin has been applied to reduce the number of virus-infected cells during the process of poultry cultivation since it can inhibit the first step in N-glycosylation (Zhu et al. 2008). The 20 co-existed classes are *macrolide-lincosamide-streptogramin* (27%/26% for mono-digestion and co-digestion respectively), *glycopeptide* (20%/19%), *multidrug* (14%/14%), *polymyxin* (12%/10%), *trimethoprim* (10%/10%), *aminoglycoside* (4%/3%), *tetracycline* (4%/5%), *mupirocin* (2%/2%). *Macrolide-lincosamide-streptogramin* (MLS) antibiotics, which consists of *erythromycin*, *clindamycin*, and *spiramycin*, are one of the most popular used antibiotics used by veterinarians in the United States. Glycopeptide antibiotics, including *vancomycin* and *teicoplanin*, have been used widely as growth promotants for livestock (Van Bambeke et al. 2004). Multidrug is also another type of antibiotic widely used in cattle and cow farming, while some bacteria can long-term persistent and result in influence mitigating animal and public health risk (Cobbold et al. 2006). *Polymyxin* is a group of polypeptide antibiotics includes *polymyxin B* and *polymyxin E*, or *colistin*, which are typically used for treating intestinal infections. Based on 2015 Summary Report on Antimicrobials Sold or Distributed for Use in Food-Producing Animals, *tetracyclines* were accounted for 62% of domestic sales, *Mupirocin* was identified to decrease the risk of infection of *methicillin-resistant staphylococcus aureus* (Lozano et al. 2011).

In addition, antibiotic resistance can be transferred by resistant genes by inactivation, protection, substitution, efflux pumping, thus a closer look was performed at the found antibiotic genes in the digestion samples. Among the two types of digestion, 188 gene groups were annotated, with 141 in mono-digestion and 172 in co-digestion, 126 were shared (Figure 2-3 and Table S1). The dominant genes in the mono digestion show similar pattern, with *vanUG* (*glycopeptide*, 16%/15% mono-digestion and co-digestion respectively), *dfrE* (*trimethoprim*, 10%/10%), *msrB* (*macrolide-lincosamide-streptogramin*, 11%/10%), *PmrE* (*polymyxin*, 11%/9%), *LlmA* (*macrolide-lincosamide-streptogramin*, 7%/9%) ranked as the high abundant ARG subtypes in both digesters (Figure 2-4). The high similarity persistence of large percentage of ARG categories and subtypes in both digesters may suggest the persistent characterization of ARGs during the anaerobic digestion. *VanUG* is a *VanG* variant resistance to vancomycin, while vancomycin resistance firstly emerged in *Enterococci* (Hayes et al. 2005), and then also found in *Staphylococcus aureus*. It is also an important part of the regulatory system of *VanG* operon, but the role of *VanUG* who encodes a transcriptional regulator in the operon process remains unknown (Depardieu et al. 2015). *dfrE* is specific resistant to trimethoprim. It is a chromosome-encoded dihydrofolate reductase, which replaces or substitutes antibiotic action target, and this process will result in antibiotic resistance. *MsrB* is an ABC-F subfamily protein that confers resistance to erythromycin and

streptogramin B antibiotics. It is located in large plasmids. *PmrE* required for the synthesis and transfer of 4-amino-4-deoxy-L-arabinose (Ara4N) to Lipid A, which allows gram-negative bacteria to resist the antimicrobial activity of cationic antimicrobial peptides and antibiotics such as polymyxin. *LlmA* is resistant to erythromycin and streptogramin B antibiotics and also associated with plasmids.

3.2 Taxonomy composition of the possible ARBs

More than 44 Gb of raw reads for each sample was produced from Illumina HiSeq sequencing. Both mono-digestion and co-digestion samples sustained a mapping ratio of greater than 68%. This result is fairly well when compared with other anaerobic digestion metagenomic data (Campanaro et al. 2016), which suggests a relatively good recovery of the reads. For discovering the potential host of the ARGs, all predicted ORFs detected with ARGs were extracted and search for the taxonomy composition to reveal the microbial diversity of the anaerobic digestion.

As a result, for all contigs contained ARGs in both digestors, bacteria were always the dominant domain, accounting for more than 98% contigs. The taxonomy affiliation at the phylum level was determined (Figure 2-5). In both digestion samples, the microbial community of ARG contained contigs were dominantly comprised of four phyla, *Firmicutes* (44.6%, 51.2% of the contigs respectively in mono-digestion and co-digestion), *Proteobacteria* (30.2% and 22.7%), *Actinobacteria* (8.8% and 11.7%) and *Bacteroidetes* (8.4% and 7.8%). However, of all the ARG contained contigs, both digesters contained a very low archaea existence, with only 1.8% and 1.2% contigs detected. Most of these affiliations are possible pathogen originated phyla. Compared with the taxonomy composition of the digesters, these microbial communities are the most abundant phylum existed in the digesters, which may suggest the ARG are more likely to choose the abundant indigenous bacteria as the carriers. It may also reveal that the indigenous bacteria communities may have stable characteristics, which are much easier to carry these long stable existed bacteria. The discovering of multiple taxonomy affiliations of ARGs in this study imply the capability of the metagenomic method. The relative consistent microbial composition between mono-digestion and co-digestion may further support the persistent of ARGs since the possible carriers' phyla seem similar.

3.3 Functional profile of the microbial communities

For functional annotation was done on the assembled into contigs, and then open reading frames (ORFs) were predicted from the contigs for saving more useful information. Then ORFs were functionally annotated in EggNOG database.

A total of 738 and 946 genes were predicted for mono-digestion and co-digestion respectively. The putative functional role for the ARG existed reads was assigned to taking into their account by the annotation of COG database. As a result, COG classes revealed that only 17 of the 25 basic metabolic categories existed in the ARG contained contigs in both types of digestion sample. Mono-digestion and co-digestion showed a similar trend in the relative distribution of metabolic categories. Cell wall/membrane/envelope biogenesis, transcription, post-translational modification, protein turnover, and chaperones, nucleotide transport and metabolism were the more harbored COG classes (Figure 2-6).

In addition, we further analyzed the functional profiles of the top significant mechanism in the digestions. These mechanisms include some transporters, resistant proteins, restriction enzymes, multidrug efflux pump (Figure 2-7). The resistant protein contains metallo-beta-lactamase domain protein, acriflavin resistance protein, multidrug resistance protein, and other resistant protein. MFS antibiotic efflux pump also exists. It includes a significant number of bacterial drug and multidrug efflux pumps, which is one of the largest groups of solute transporters (Lekshmi et al. 2018). Some transcriptional regulators modulate antibiotic resistance and have the function of quorum sensing (Dieppois et al. 2012). Acetyltransferase enzyme confers resistance to some aminoglycoside resistances, such as kanamycin, gentamicin, and tobramycin. The most abundant function thymidylate synthase could be associated with the drug resistance (Yeh et al. 1998). The existence of rRNA adenine N-6-methyltransferase protein is a Macrolide-lincosamide-streptogramin B resistance protein, which

can reduce the affinity between ribosomes and macrolide-lincosamide-streptogramin B antibiotics (Renaux and Consortium 2018). Aminoglycoside-2"-adenylyltransferase (AAD(2")) is coded by a plasmid-borne gene *aadB*, which mediates gentamicin, kanamycin, and tobramycin resistance phenotype in bacteria. Gentamicin, kanamycin and tobramycin resistances belongs to aminoglycoside, and some aminoglycoside resistances were found in both digesters. SAM (S-adenosyl-L-methionine) is an enzyme, is the source of the appended group in a majority of methylation reactions (Bauerle et al. 2015), commonly regarded as SAM-dependent methyltransferases, could confer resistance to a range of ribosome-targeting antibiotics (Stojkovic et al. 2017). The relative abundant existence of predicted proteins contain in different types of antibiotic response may suggest that anaerobic digestion has the possibility of harboring a diversity of ARGs, and the co-abundant of both digestions may imply the possible similar trend of ARGs existence.

3.4 Co-occurrence of ARGs

For identifying which bacteria were carrying the ARG genes, the network analysis was used in revealing the co-occurrence pattern between ARG subtypes and the microbial community. Out of 110 potential hosts in both digesters, 55 of them were shared among both types of digesters. It is interesting to find that multiple potential hosts were found in some ARGs in both digesters (Figure 2-8). For example, *LlmA* and *msrB* from multidrug were found to be affiliated with more than ten bacteria. *LlmA* was found linked to *Butyrivibrio*, *Lachnospira*, *Paenibacillus*, *Peptoclostridium*, *Ruminiclostridium*, *Ruminococcus*, etc., while *msrB* was associated with *Acinetobacter*, *Clostridium*, *Cystobacterineae*, *Methanothrix*, *Pseudomonas*. Two subtypes from glycopeptide, *VanUG* was found affiliated with *Acidaminococcus*, *Clostridium*, *Holdemanella*, *Lachnoclostridium*, *Lachnospiraceae*, *Alkaliphilus*, *Desulfotomaculum*, *Dialister*, *Eubacterium*, *Levyella*, *Mitsuokella*, *Oribacterium*, *Selenomonas*. *dfrE* was inconsistent affiliated with *Arcanobacterium*, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Carnobacterium*, *Enterococcus*, *Knoellia*, *Lactobacillus*, *Prevotella*, *Proteiniphilum*, *Thauera*. In addition, some bacteria were found to have multiple potential resistance during the anaerobic digestion. *Clostridium* potentially carry four ARG subtypes (*lnuD*, *msrB*, *vanU*, *VanUG*). *Lachnospiraceae* may carry four ARGs (*VanUG*, *LlmA*, *LlmA*, *PmrE*). *Pseudomonas* possibly take five subtypes (*arpC*, *cpxr*, *golS*, *KsgA*, *msrB*), and *Ruminococcus* could have three subtypes (*LlmA*, *PmrE*, *vanWG*). Such inconsistent existent of ARGs and its potential hosts may imply the survival environment and the selection pressure in the digesters give rise to the co-existent of pathogens and bacteria.

It has been mentioned that anaerobic digestion could reduce the ARGs abundance since anaerobic digestion can remove some pathogens and DNA hydrolysis. However, it is interesting to find that in this study most ARGs consistent exist even after the co-digestion, which suggested a mixed effect of anaerobic digestion on ARGs. The percentage difference and composition difference in ARGs might be attributed to the changes in the microbial community structure. Because co-digestion cultured a group of different microbial community than mono-digestion, and these microbial community consequently harbored these genes and mobile elements. However, future studies are needed to validate the correlation between ARGs and its potential hosts.

The co-occurrence of Mobile Genetic Elements (MGEs), consist plasmids, transposase, recombinase, transposon, conjugative, integrase, and conjugal. MGEs play significant roles in the mobility and acquisition of ARG among a diverse of microbial community by the horizontal gene transfer pathway. Therefore, ORFs were searched against the NCBI retrieved MGE-associated sequences.

Under an identity of 50, coverage 50, e-value < 1e-10, probability > 50%, 591 and 780 ORFs were identified with MGEs. Both types of digester were inconsistently found with plasmids and transposase (Figure 2-9). Plasmids were found as the most abundant MGEs, while transposase and recombinase were also found in the samples. Plasmids have been regarded as the most important part for intercellular horizontal transfer of genetic materials (Zhang et al. 2011), the presence of plasmid indicates that the microorganism in the samples have the gene-mobilizing capacity (Garbisu et al. 2018). It has previously been identified that ARGs with plasmids were more likely to transfer horizontally by conjugation (Becker et al. 2018). Transposase is an enzyme that binds to the end of a transposon and catalyzes its movement to another part of the genome by a cut and paste mechanism or a replicative transposition mechanism. The potential for horizontal transfer of ARGs because

of transposon-specific ARGs is implicated by the enrichment of transposases, and there could be a high correlation between ARG and transposase abundance (Zhu et al. 2013).

3.5 Persistent of potential antibiotic-resistant bacteria

For further investigate whether the persistence characteristic existed in the digesters, samples produced after around three years of the digesters were sequenced by 16S rRNA sequencing. After denoised the sequences, the raw 16S sequences were extracted and analyzed with NCBI RefSeq database used in metagenomic data. 16S reads and assembled contig that can be mapped to the same genome reads were regarded as the same origin. As a result, some potential antibiotic-resistant bacteria were still found in the digesters, which suggests that some bacteria may have long-term stable characteristics and reluctant to change even under the anaerobic environment. In details, similar to the former ARBs, a variety of bacteria mainly from *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* were found inconsistent existed in the digesters. Besides the sequences cannot be classified from the genus level, *Proteiniphilum*, *Paenibacillus*, *Bacteroides*, *Ruminiclostridium*, *Syntrophus*, *Streptococcus*, *Clostridium*, *Bacillus*, etc. were the genus found consistent in the digesters. These anaerobes are typically found in the anaerobic digestion environment, while some of them, such as *Bacteroides* and *Clostridium*, have longed been studied as important ARBs (Gajdacs et al. 2017). The co-existence of these bacteria may imply the consistent prevalence of some bacteria, and these bacteria may provide opportunities for the survival of the ARGs.

The persistence of ARGs could be owed to various selective forces and mechanisms, such as selective pressure, beneficial fitness resistance, plasmid-mediated increase in fitness, etc. Two types of digester have been fed with manure for a long time, while the central microbial communities will not change once they achieve a stable status. The long-term performance of the digesters has formed a stable selective pressure of the digesters. The sudden addition of poultry waste will not increase the leading functional microbial community will increase during the treatment and will recovery during the post-digestion. Thus, there may a niche microbial community existed in the digesters. Previous studies have found that even without the addition of antibiotics, the persistence of high frequency of resistant still can be found (Barbosa and Levy 2000). Therefore, a sudden disturb of the self-recovery ability contained digesters may support the persistence of ARBs. The unchanged ARBs may provide a similar plasmid mediation function, and the ARGs could persist.

4. Conclusion

For checking the persistence of ARGs and its corresponding potential hosts, the metagenomic approach was applied to the mono-digestion of manure samples and the co-digestion of poultry waste and manure samples. Under a relative good assembly of the short reads into long contigs, contigs were then transformed into the open reading frames. Taxonomy was then assigned and annotated functional profiles. This may provide a better understanding of the samples.

As a result, compared with mono-digestion, a variety of ARGs remained in the digesters after co-digestion, mainly expressed in *macrolide-lincosamide-streptogramin*, *glycopeptide*, and *multidrug* ARG types, 126 unchanged subtypes. The most abundant unchanged subtypes were *vanUG*, *dfrE*, *msrB*, *PmrE*, *LlmA*. Such persistence of ARGs was accompanied by the existence of some bacteria, which mainly originated from *Firmicutes*, *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes*.

Furthermore, the functional profiles of the microbial community in both digesters were identified to be identical in both digesters. After annotating the ORFs to the EggNOG database, mechanisms include some transporters, resistant proteins, restriction enzymes, multidrug efflux pump were found commonly existed.

In addition, multiple potential hosts were found in some ARGs in both digesters, while some bacteria were found to have multiple potential resistance during the anaerobic digestion. Besides, MGEs were also found similar in both digestions. This phenomenon was found in both digesters, which may suggest the existence of potential host unchangeable.

Finally, a closer look at the microbial community after the long-term history of the digesters was checked by 16S rRNA approach. It was found that there indeed exist some bacteria stable and long-term existed in both digesters, which may further support the persistence of ARGs and potential hosts in the digesters.

Appendix B: Figures and Tables

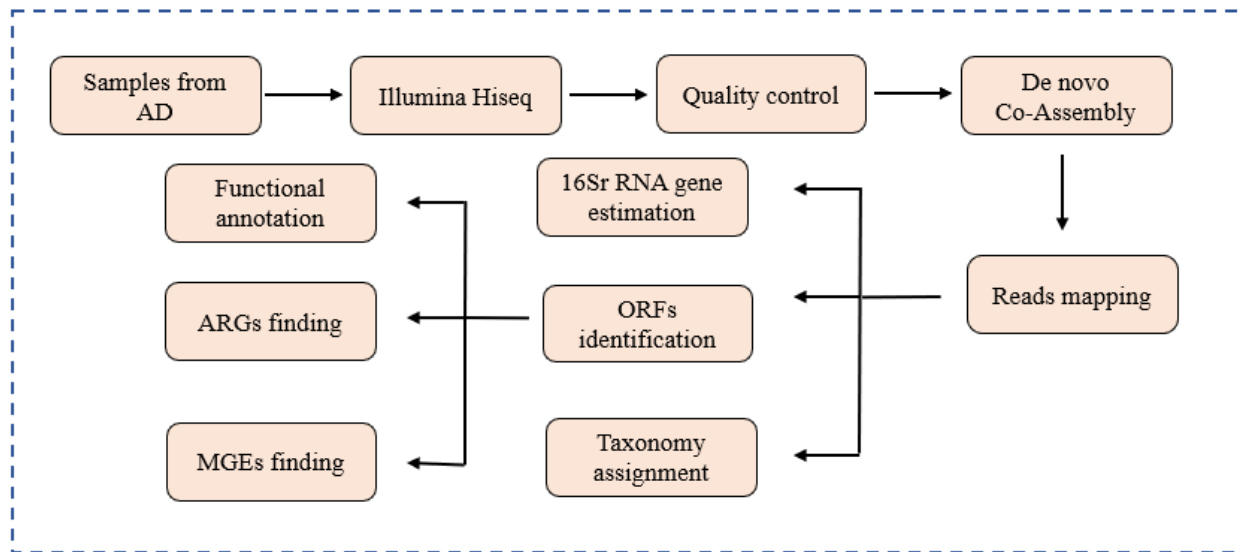


Figure 2-1 Analysis workflow for ARGs identification on metagenomic data.

Table 2-1 Summary of sample contribution to the assembly.

Sample ID	#Raw reads	#Reads after QC	#Assembled contigs	N50	N75
C	1,172,049,040	1,026,753,626	207652	7204	3516
D	1,132,164,528	1,038,646,338	230337	7715	3631
Sample ID	Total length ($\geq 2000\text{bp}$)	GC content	Reads mapped	Mapping ratio	
C	1,175,483,181	51.19%	711,682,215	68.43%	
D	1,348,705,567	50.68%	711,715,607	68.52%	

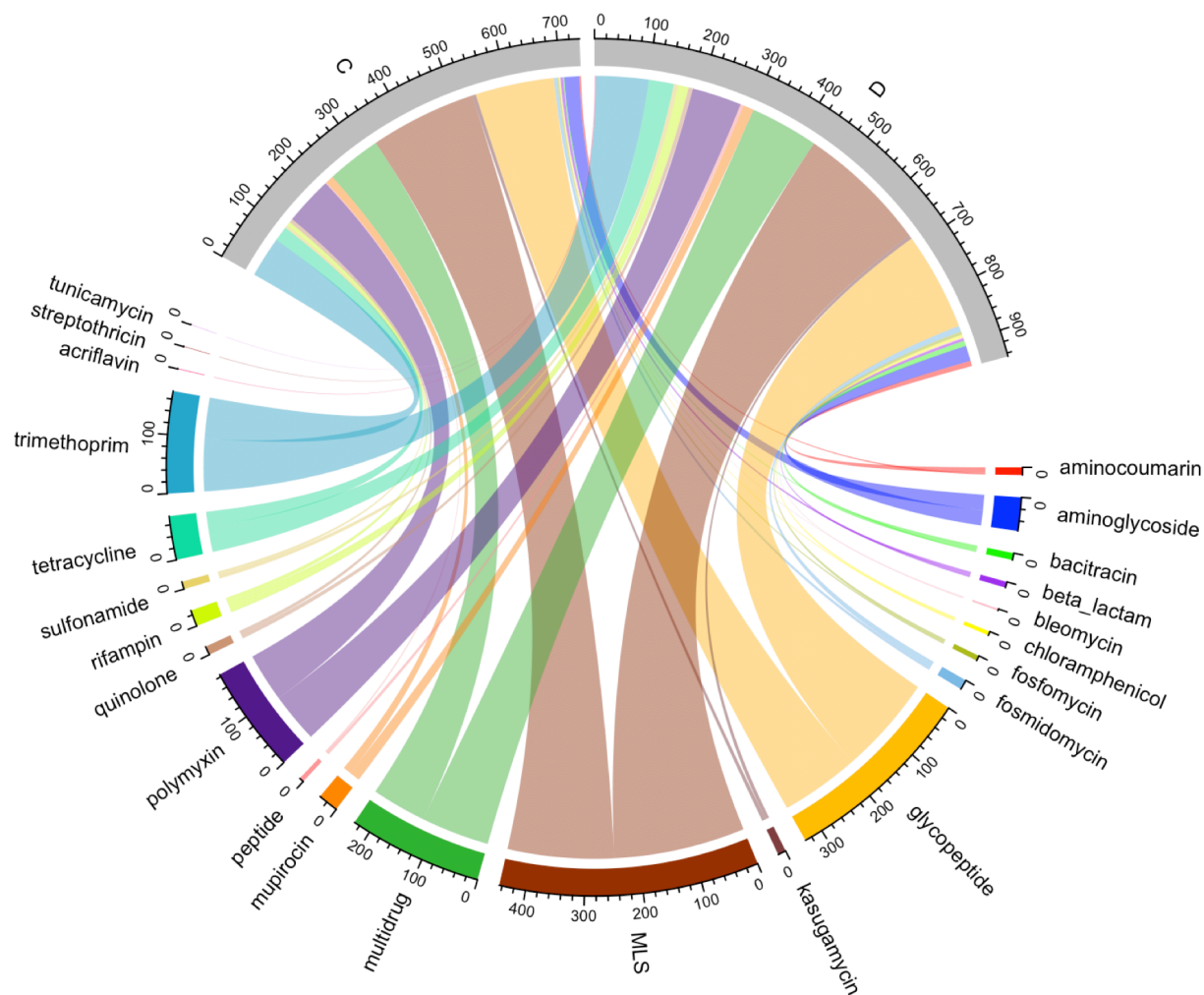


Figure 2-2 Distributions of ARG types detected in digesters.

MLS denotes macrolide-lincosamide-streptogramin. The outmost circle represents the gene number of ARG types. The length of the bars on the outer ring represented ARG gene found in two digesters (grey bars of the diagram, labeled as C and D) and correlates the gene number of respective ARG types in the digesters (colorful bars). Each color represents an ARG type.

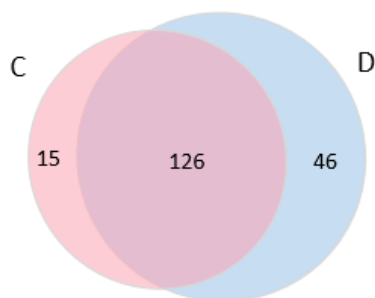


Figure 2-3 Venn diagram for the ARGs detected among the two treatments.

Unique ARG subtypes in C digester is 142, while in D digester is 173. They have 127 ARG subtypes in common, and some unique in their own. C represents digester C, while D is digester D.

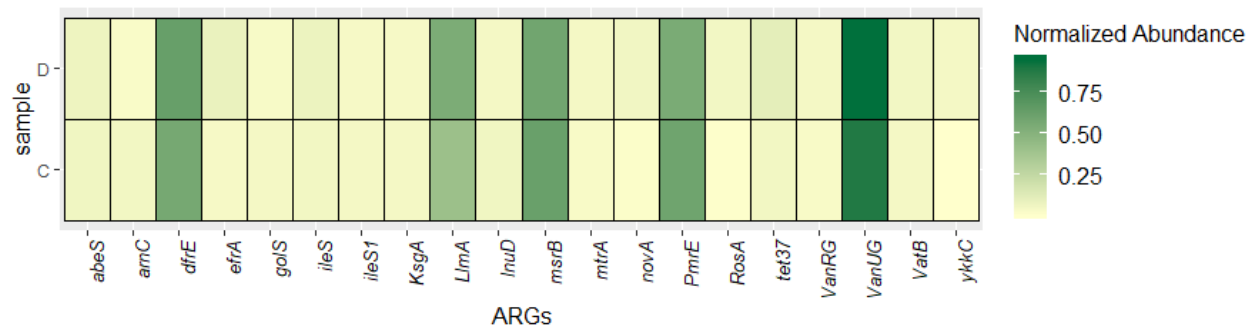


Figure 2-4 The abundance of the ARGs subtypes in digester samples.

The color intensity in each panel shows the ARGs/16s rRNA gene copy number. For the visualization size limit, normalized abundance greater than 0.2 were displayed.

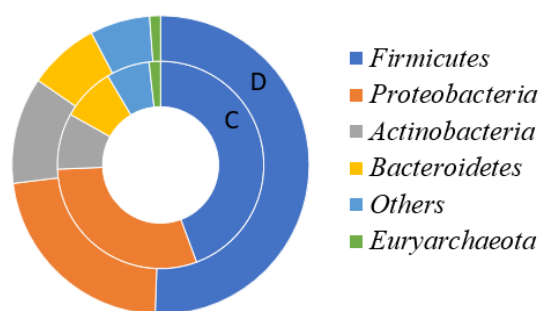


Figure 2-5 Relative consistent microbial community among the ARG carried contigs at the phylum level.

The inner circle represents co-digestion samples, while the outer circle represents mono digestion sample.

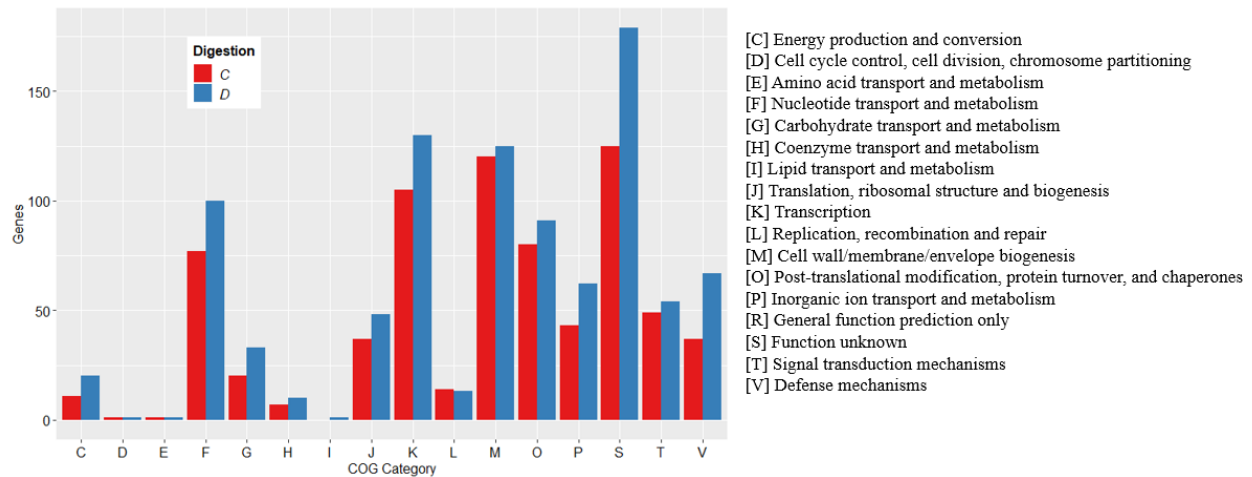


Figure 2-6 Functional class of both samples in digester C and D.

Gene contains in each virus annotated assembled contigs were predicted by Prodigal. Genes were predicted and annotated against the eggNOG database with $e\text{-value} < 10^{-5}$. The values shown in this figure is the mean value of each treatment group. Selection of one representative sample for each treatment was displayed.

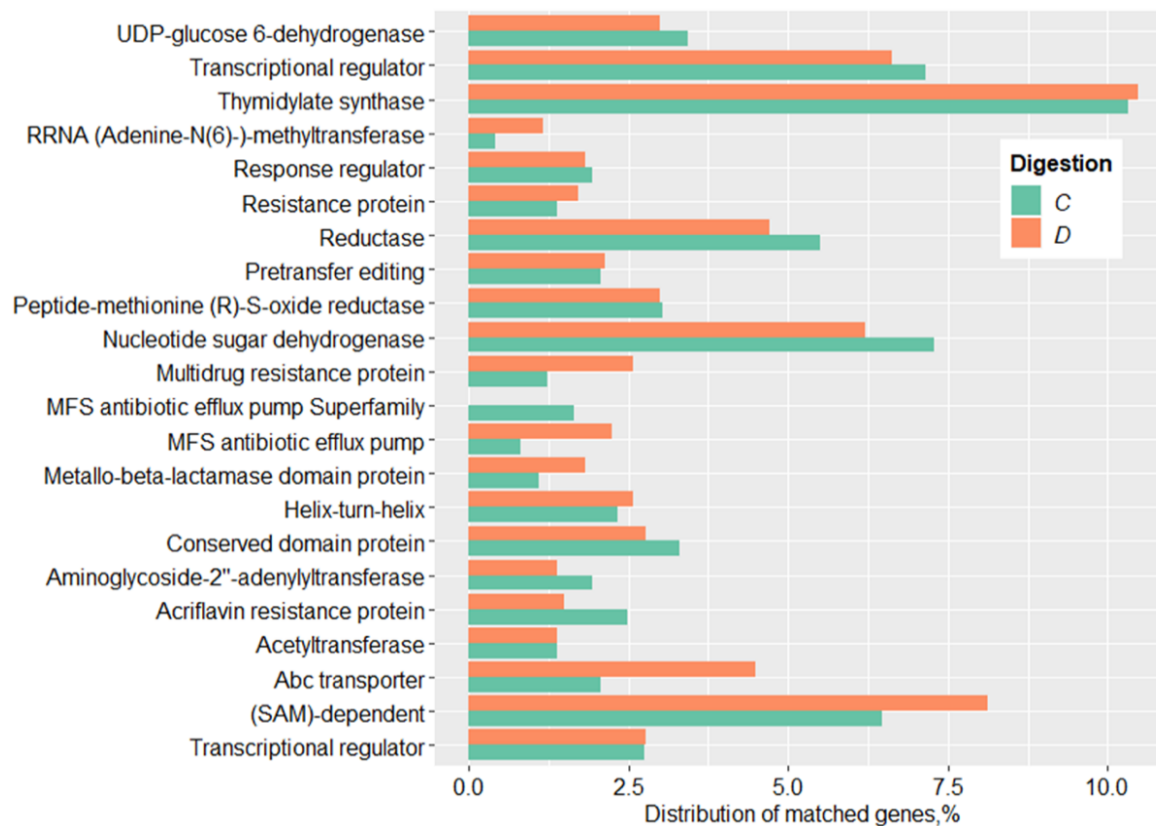


Figure 2-7 Functional class of defense mechanisms in digester C and D.

Metagenomic data were annotated against the eggNOG database at a cutoff of E-value $< 10^{-5}$

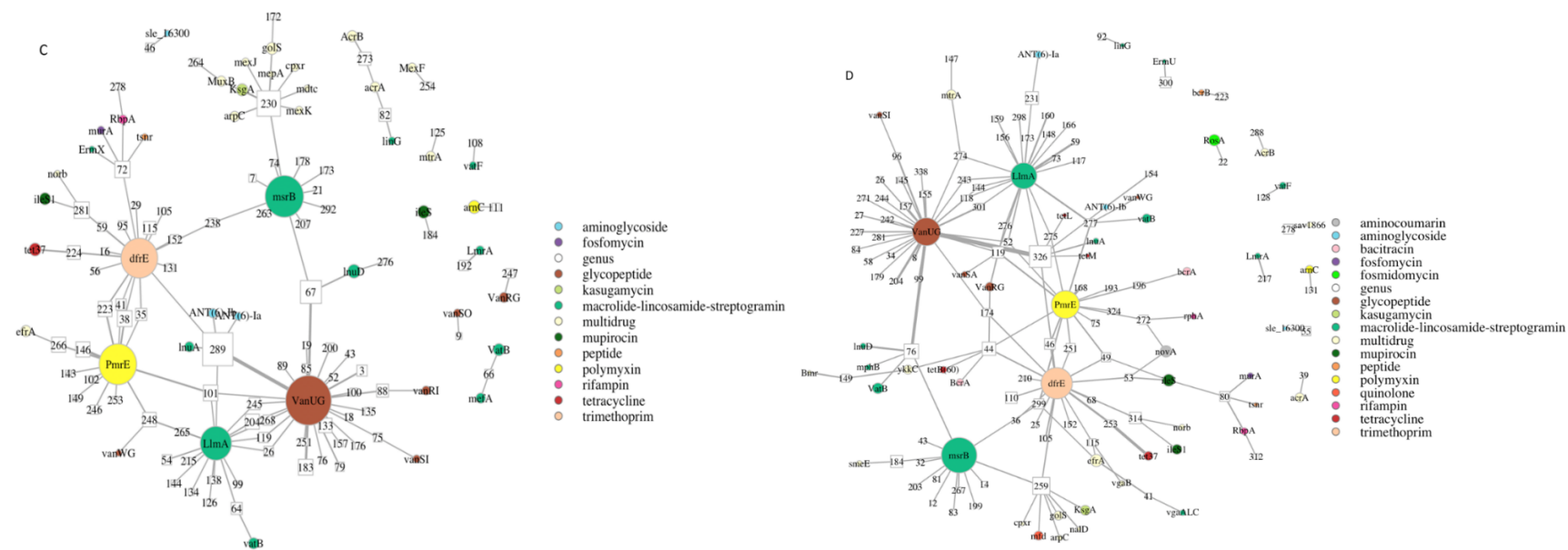


Figure 2-8 Co-occurrence of ARGs and potential taxonomical origin (ARBs).

The colorful circle nodes stand for various ARG groups (each color stands for one ARG category), while the white squared nodes indicate diverse microorganisms, labeled as ‘genus’ in the legend. Microorganism nodes listed were all at genus/species/strain level. The node size shows the connection of ARGs and microorganisms. C stands for mon-digestion, while D stands for co-digestion.

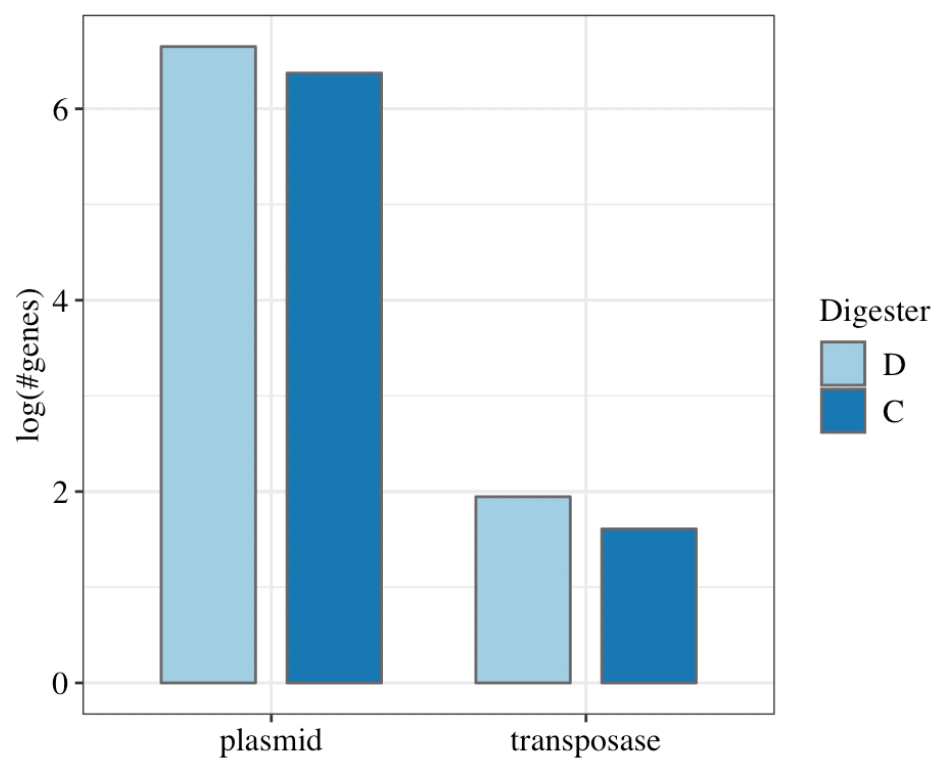


Figure 2-9 Summary of MGE abundances found in the metagenome samples.

MGEs were identified by screening the ORFs by a set of MGEs from NCBI. The abundance was calculated by taking a log of the matching ORFs. Digester D represents the co-digestion, while digester C is mono-digestion.

Table S1 Taxonomy of the microorganisms in the networks.

genus	nodes_C	genus	nodes_D
<i>Acidaminococcus</i>	3	<i>Acidaminococcus</i>	8
<i>Acinetobacter</i>	7	<i>Acidobacterium</i>	12
<i>Actinoplanes</i>	9	<i>Acinetobacter</i>	14
<i>Alistipes</i>	16	<i>Aliihoeflea</i>	22
<i>Alkaliphilus</i>	18	<i>Alkaliflexus</i>	25
<i>Allofustis</i>	19	<i>Alkaliphilus</i>	26
<i>Aminobacterium</i>	21	<i>Allofustis</i>	27
<i>Anaerotruncus</i>	26	<i>Anaerococcus</i>	32
<i>Arcanobacterium</i>	29	<i>Anaerotruncus</i>	34
<i>Bacillus</i>	35	<i>Arcanobacterium</i>	36
<i>Bacteroides</i>	38	<i>Aromatoleum</i>	39
<i>Bifidobacterium</i>	41	<i>Atopococcus</i>	41
<i>Blautia</i>	43	<i>Azospirillum</i>	43
<i>Brevibacterium</i>	46	<i>Bacillus</i>	44
<i>butyrate-producing bacterium SS3/4</i>	52	<i>Bacteroides</i>	46
<i>Butyrivibrio</i>	54	<i>Bifidobacterium</i>	49
<i>Candidatus Omnitrophus</i>	56	<i>Blautia</i>	52
<i>Carnobacterium</i>	59	<i>Brachybacterium</i>	53
<i>Clostridiales bacterium NK3B98</i>	64	<i>Brevibacterium</i>	55
<i>Clostridiisalibacter</i>	66	<i>butyrate-producing bacterium SS3/4</i>	58
<i>Clostridium</i>	67	<i>Butyrivibrio</i>	59
<i>Corynebacterium</i>	72	<i>Carnobacterium</i>	68
<i>Cystobacterineae</i>	74	<i>Clostridiales bacterium NK3B98</i>	73
<i>Dehalobacter</i>	75	<i>Clostridiisalibacter</i>	75
<i>Dehalogenimonas</i>	76	<i>Clostridium</i>	76
<i>Desulfitobacterium</i>	79	<i>Corynebacterium</i>	80
<i>Desulfococcus</i>	82	<i>Cutibacterium</i>	81
<i>Desulfosporosinus</i>	85	<i>Cystobacterineae</i>	83
<i>Desulfotomaculum</i>	88	<i>Dehalobacter</i>	84
<i>Dialister</i>	89	<i>Desulfococcus</i>	92
<i>Enterococcus</i>	95	<i>Desulfotomaculum</i>	96
<i>Erysipelotrichaceae bacterium NK3D112</i>	99	<i>Dialister</i>	99
<i>Ethanoligenens</i>	100	<i>Dysgonomonas</i>	105
<i>Eubacterium</i>	101	<i>Enterococcus</i>	110
<i>Faecalibacterium</i>	102	<i>Erysipelothrix</i>	115
<i>Flavobacteriaceae</i>	105	<i>Erysipelotrichaceae bacterium NK3D112</i>	117
<i>Gammaproteobacteria bacterium MFB021</i>	108	<i>Ethanoligenens</i>	118
<i>Geobacter</i>	111	<i>Eubacterium</i>	119
<i>Halomonas</i>	115	<i>Gammaproteobacteria bacterium</i>	128
<i>Holdemanella</i>	119	<i>Geobacter</i>	131

Table S1 continued

genus	nodes_C	genus	nodes_D
<i>Ilumatobacter</i>	125	<i>Holdemanella</i>	144
<i>Intestinimonas</i>	126	<i>Holdemania</i>	145
<i>Knoellia</i>	131	<i>Ilumatobacter</i>	147
<i>Lachnoclostridium</i>	133	<i>Intestinimonas</i>	148
<i>Lachnospira</i>	134	<i>Jeotgalicoccus</i>	149
<i>Lachnospiraceae</i>	135	<i>Knoellia</i>	152
<i>Lachnospiraceae bacterium 28-4</i>	138	<i>Kurthia</i>	154
<i>Lachnospiraceae bacterium AC2029</i>	143	<i>Lachnoclostridium</i>	155
<i>Lachnospiraceae bacterium C6A11</i>	144	<i>Lachnospira</i>	156
<i>Lachnospiraceae bacterium FE2018</i>	146	<i>Lachnospiraceae</i>	157
<i>Lachnospiraceae bacterium P6B14</i>	149	<i>Lachnospiraceae bacterium 10-1</i>	159
<i>Lactobacillus</i>	152	<i>Lachnospiraceae bacterium 28-4</i>	160
<i>Levyella</i>	157	<i>Lachnospiraceae bacterium C6A11</i>	166
<i>Mesorhizobium</i>	172	<i>Lachnospiraceae bacterium FE2018</i>	168
<i>Methanobacterium</i>	173	<i>Lachnospiraceae bacterium YSB2008</i>	173
<i>Methanomassiliicoccus</i>	176	<i>Lactobacillus</i>	174
<i>Methanotherix</i>	178	<i>Levyella</i>	179
<i>Mitsuokella</i>	183	<i>Lysobacter</i>	184
<i>Mobilicoccus</i>	184	<i>Marvinbryantia</i>	193
<i>Nocardia</i>	192	<i>Methanobrevibacter</i>	196
<i>Oribacterium</i>	200	<i>Methanotherix</i>	199
<i>Paenibacillus</i>	204	<i>Microbacterium</i>	203
<i>Paracoccus</i>	207	<i>Mitsuokella</i>	204
<i>Peptoclostridium</i>	215	<i>Mycoplasma</i>	210
<i>Prevotella</i>	223	<i>Nocardia</i>	217
<i>Proteiniphilum</i>	224	<i>Oceanobacillus</i>	223
<i>Pseudomonas</i>	230	<i>Oribacterium</i>	227
<i>Rhizobium/Agrobacterium group</i>	238	<i>Paenibacillus</i>	231
<i>Ruminiclostridium</i>	245	<i>Peptoanaerobacter</i>	242
<i>Ruminococcaceae bacterium AB4001</i>	246	<i>Peptoclostridium</i>	243
<i>Ruminococcaceae bacterium D16</i>	247	<i>Peptoniphilaceae</i>	244
<i>Ruminococcus</i>	248	<i>Prevotella</i>	251
<i>Selenomonas</i>	251	<i>Proteiniphilum</i>	253
<i>Sideroxydans</i>	253	<i>Pseudomonas</i>	259
<i>Singulisphaera</i>	254	<i>Rhizobium/Agrobacterium group</i>	267
<i>Staphylococcus</i>	263	<i>Roseburia</i>	271
<i>Stenotrophomonas</i>	264	<i>Ruania</i>	272
<i>Stomatobaculum</i>	265	<i>Ruminiclostridium</i>	274
<i>Streptococcus</i>	266	<i>Ruminococcaceae bacterium AB4001</i>	275
<i>Subdoligranulum</i>	268	<i>Ruminococcaceae bacterium D16</i>	276

Table S1 continued

genus	nodes_C	genus	nodes_D
<i>Syntrophorhabdus</i>	273	<i>Ruminococcus</i>	277
<i>Tepidanaerobacter</i>	276	<i>Salinicoccus</i>	278
<i>Tetrasphaera</i>	278	<i>Selenomonas</i>	281
<i>Thauera</i>	281	<i>Smithella</i>	288
<i>unknown</i>	289	<i>Stomatobaculum</i>	298
<i>Vibrio</i>	292	<i>Streptococcus</i>	299
		<i>Streptomyces</i>	300
		<i>Subdoligranulum</i>	301
		<i>Tetrasphaera</i>	312
		<i>Thauera</i>	314
		<i>Treponema</i>	324
		<i>unknown</i>	326
		<i>Youngiibacter</i>	338

CHAPTER 3
POLLUTANTS MITIGATION IN ENGINEERED PROCESS

Section A: Legacy Effects Influence the Pathway Selection in Anaerobic Digestion

Abstract

To determine adaptive changes in the methanogenic populations in response to substrate spiking, and to identify potential linkages between process performance and microbial community, we modeled anaerobic digestion process with triplicate continuous anaerobic digesters. The process was subsequently disrupted by formate loading. 16S rRNA sequence analysis of microbial communities responding to formate overloading showed that populations surprisingly related to *Methanosaeta* were the dominant methanogens before and after formate spiking, suggesting the functional importance of these acetoclastic methanogens, not hydrogenotrophic methanogens, in balanced anaerobic digestion processes. Population variation in possible homoacetogens fluctuated with *Methanosaeta*. Those homoacetogens likely helped *Methanosaeta* convert hydrogenotrophic methanogenesis function to the acetoclastic function, which was further supported by K-means clustering analysis. Results show that highly ranked population with low-rank variance were ‘home’ population with legacy effects. The “home” bacteria and “home” methanogen coexisted and simultaneously worked together, which could lead to the pathway redirection in the formate treatment. Understanding the connections between microbial populations and the process performance discovered in this study is benefits for strategies development in the mitigation of the impact of substrate loading on anaerobic digestion processes. Further studies are required to identify mechanisms contributing to the unexpected pathway redirection during formate treatment observed in this study.

Keywords: Formate; legacy effects; acetoclastic methanogenesis; hydrogenotrophic methanogenesis; homoacetogens; sequencing

1. Introduction

The biological process has been widely known as an essential part of wastewater treatment. Different processes, such as biological treatment, activated sludge, and anaerobic digestion, are included. Anaerobic digestion has been recognized as the unneglectable treatment process, which profoundly influences the treatment efficiency by the bacterial metabolism (Cyzdik-Kwiatkowska and Zielińska 2016). Therefore, understanding the interactions of the microbial community is required to develop strategies for an active process.

The wastes, which contains abundant protein, carbohydrates, and lipids (Angelidaki and Sanders 2004, Liu and Whitman 2008), could be biologically transformed into methane. Extensive studies have been focused on a variety of source wastes involved in the transformation, which refer to animal waste (Chen et al. 2017, Wijesinghe et al. 2017, Wu et al. 2016a, Zhang et al. 2017b, Zhang et al. 2017c), straws (Mao et al. 2017, Romero-Güiza et al. 2017), food waste (Berry et al. 2017, Maragkaki et al. 2018, Palatsi et al. 2011, Svensson et al. 2018), etc. Different bacteria can thrive through fermentation reactions, soluble organic compounds were released and then consumed by other bacteria. This biological conversion is the well-studied chain of anaerobic processes, which includes hydrolysis, acidogenesis, acetogenesis, and methanogenesis. The process chain is highly sensitive, therefore its efficiency was significantly influenced even by minor change of operational conditions, such as temperature (De Vrieze et al. 2015, Lin et al. 2017, Westerholm et al. 2017, Xiao et al. 2018), pH (Latif et al. 2015, Wang et al. 2014), ammonia concentration (De Vrieze et al. 2015), hydraulic retention time (Dareioti and Kornaros 2015, Ziganshin et al. 2016), organic loading rate (Wirth et al. 2015), reactor scale, moisture content. Different functional bacteria and archaea are essential participants in these processes. Cellulolytic and other polymer degrading bacteria are active in transforming complex polymers into monomers (Angelidaki and Sanders 2004, Batstone et al. 2002, Gujer and Zehnder 1983, Mata-Alvarez et al. 2000). Fermentation bacteria involved in the primary fermentation, in this acidogenesis process short-chain fatty acids, alcohols, acetate, C1 compounds, H₂ and CO₂ are produced (Myint et al. 2007). Methanogenic archaea help produce methane. For understanding complex microbial community structures in AD processes under transient conditions, considerable efforts have been devoted.

Even the complexity of the system, final substrates utilized by the methanogens only included three types, short-chain fatty acid (C1–C6), n- or i-alcohols and gases (CO, CO₂, and H₂) (Zhang et al. 2014). The critical functional methanogens include two parts, acetoclastic methanogens that mainly use acetate, and hydrogenotrophic methanogens that mainly use H₂/CO₂ or formate. Hydrogenotrophic methanogens are more resistant than acetoclastic methanogen especially under the harsh operating environment (Yenigün and Demirel 2013). The proper utilization of the hydrogenotrophic pathway under harsh environment could be an effective way to improve the performance of anaerobic digestion. A deeper understanding of the hydrogenotrophic methanogenesis procedure should be a valuable addition to our knowledge of the microbial process.

Formate and hydrogen are crucial electron carriers in the final process of methanogenesis (Boone et al. 1989, Pan et al. 2016, Schink et al. 2017), especially the hydrogenotrophic methanogenesis. They determine the energy balance of the facultative and obligate syntrophic methanogenic community. Anaerobic bacteria are much more energetic advantageous than methanogens in facultative syntrophy process, while bacteria and archaea mutually metabolize on one substrate in obligate syntrophy. Hydrogen is continuously produced and utilized by hydrogenases microorganisms, while the dehydrogenases genes produce formate. Previous literature has given much emphasis on the hydrogen, but little attention has been paid to formate. Formate has already been detected in simultaneously work with hydrogen in the anaerobic system. Moreover, former diffusion model has suggested that formate can sustain a 100-fold higher conservation rate than hydrogen (Stams and Plugge 2009). For this reason, more efforts should be put to understand the microbial community structure in anaerobic formate treatment.

In addition, there could be mutually transformation between the two well-known methanogenesis pathway, the conversion between formate and acetate. Formate can be indirectly transferred into acetate by homoacetogenesis pathway (Nie et al. 2007, Rachbauer et al. 2017, Ryan et al. 2008, Saady 2013), a pathway being considered thermodynamically unfavorable under natural environment. Therefore, the homoacetogens-occupied pathway was thought to be more energy consumption and could not outcompete methanogens for a common substrate, H₂. Homoacetogenesis was favorable under low H₂ partial pressure (Goodwin and Zeikus 1987, Liu et al. 2016),

low temperature (Conrad and Wetter 1990, Kotsyurbenko 2005, Kotsyurbenko et al. 2001), acidic environments (Phelps and Zeikus 1984). Homoacetogens could grow faster than hydrogenotrophic methanogens when under a hydrogen threshold. Homoacetogenesis has also been identified to function with some homoacetogens, *Clostridium*, *Clostridia*, *Acetobacterium* strains (Esquivel-Elizondo et al. 2016, HEISE et al. 1992, Levin et al. 2004). However, under what condition can make this conversion happens is still unknown. Currently, limited research has been focused on determining how this single substrate change affect the metabolism of the microbial community.

Digester configuration is a suggested way to capture biogas during organic waste storage, which can be a substitute for other non-regeneration energy and a grateful way to treat organic waste. Reactor configuration has also been identified as an essential factor that can profoundly influence the biogas production efficiency. Different reactors, like conventional anaerobic reactors, sludge retention reactors and anaerobic membrane reactors have been studied for a long time. The continuous batch reactor is of high reliability, and it has been widely used for wastewater that contains the high strength of livestock wastewater, industrial organic waste (Mao et al. 2015a, South et al. 1995). Its property of operability and controllability make it a great modeling tool in the lab. Stable operation performance of CH₄ production was regarded as the most critical indicator in reactor evaluation. Experimental study on short-term reactor set-up was the most popular one, which was considered as understandable and reasonable. The complex relationship during the short treatment period in the anaerobic food web has been widely studied. However, little attention has been paid to the influence of the starting environment of the reactors on the microbial community. Previous literature has listed the successional change of the microbial community occurred even under no disturbance of the environment (Wu et al. 2016a). However, whether the successional pattern can influence the later treatment remains still unclear.

The objective of this study was to investigate what population metabolizes formate in the anaerobic digestion system with a long history. Triplicate continuous bench scale reactors and batch reactors were configured. We manipulated the contemporary environment by two-time period spiking of formate. The microbial community structure in different treatment periods was examined by Illumina Miseq sequencing of the 16S rRNA gene. It is interesting to find that hydrogenotrophic methanogens were not the corresponding key archaea during formate treatment, which may indicate the impact of legacy effects. We expect this could help understanding more about the formate role in the anaerobic digestion, further understanding the methanogenic process among the microbial communities. The application of legacy effects may help to disentangle the ecology of the anaerobic food web.

2. Materials and Methods

2.1 Dilution medium

The basal medium was prepared according to the following recipe (per liter) as previously described (He and Sanford 2002): NaCl, 1.0 g; $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 0.5 g; KH_2PO_4 , 0.2 g; NH_4Cl , 0.3 g; KCl, 0.3 g; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.015 g; trace element solution, 1.0 mL; Se/Wo solution, 1.0 mL; and resazurin, 1.0 mg. L-cysteine (0.031 g/L) and $\text{Na}_2\text{S} \cdot 9\text{H}_2\text{O}$ (0.048 g/L) were added as reductants into the basal medium after it had been boiled and cooled to room temperature under an oxygen-free N_2 atmosphere. NaHCO_3 (2.52 g/L) was added to the medium as the buffer. A sterile vitamin solution (1%) was added after autoclaving (Wolin et al. 1963).

2.2 Substrates and inoculum

To study the responsive methanogenic populations, formate was daily fed into the continuous anaerobic reactors, which have been operated for more than nine years. The feedings were separated into two periods of spiking, each spiking period is composed of 10mM formate respectively for two days, and 20mM formate respectively for five days.

To demonstrate the influences of formate treatment on methanogenesis pathway selection. Batch reactors were newly set up with an initial anaerobic basal medium, and a 10% (vol/vol) inoculum of the same digestate from the continuous anaerobic digesters, flushed with pure N_2 and sealed with butyl rubber stoppers and aluminum caps. Formate was also added as the sole substrate when the formate was depleted, as indicated by no methane production. The inoculum was also transferred from the continuous reactors. The second enrichment was continued by transferring a 10% (vol/vol) inoculum from the first enrichment cultures into fresh medium followed by repeated feedings of formate.

2.3 Reactors set up

Triplicate mesophilic continuous anaerobic digesters were set up prior to this study as former described (Chen et al. 2016b). The hydraulic retention time was maintained at 20 days and was sustained at a stable condition by feeding $1.0 \text{ g VS L}^{-1}\text{day}^{-1}$ from livestock wastewater loading. The performance was monitored of some environmental parameters, such as pH, VFA, methane production.

In triplicate, mesophilic batch reactors were also set up for formate enrichment study. The headspace of the reactors was sampled to analyze the gas composition. Daily measurement of gas in the headspace was identified with respect to CH_4 , CO_2 , and N_2 . Water displacement method was applied to measure the biogas production (Zhu et al. 2011). Samples were collected with a 1 mL syringe. Hewlett Packard 5890 Series II gas chromatograph (GC) equipped with a thermal conductivity detector (TCD) and a Supelco packing column (60/80 Carbonxen®- 1000; Sigma-Aldrich, St Louis, MO, USA) was applied to determine the proportion of methane content in biogas. The carrier gas was argon and the flow rate maintained at 5 mL/min.

Biomass samples were collected from the digesters periodically, which resulted to 15 samples from formate treatment, and 1 sample from the formate-enriched batch reactors. The storage temperature was maintained at -80°C before DNA extraction.

2.4 Sequencing processing of 16S rRNA gene

DNA extractions were performed using MP FastDNATM Spin Kit for Soil according to the manufacturer's instruction (Protocol Revision #116560200-201608). Briefly, the biomass samples were resuspended in the 978 μL Sodium Phosphate Buffer solution and 122 μL MT Buffer, then followed by 100 seconds vortex on MP biomedical fast prep instrument. Then the DNA was purified with Zymo Genomic DNA Clean & ConcentratorTM-10 kit according to the instructions provided by the manufacturer. The DNA concentrations were determined using Thermo Scientific NanoDrop ND-3300 Fluorospectrometer.

Polymerase chain reactions were prepared with a cocktail mix containing 12.5 μ L Phusion flash Master Mix, 10 μ L ultra-pure water, 1 μ L forward primer, 1 μ L reverse primer, and 2 μ L (100 to 150 ng) DNA template. Each universal primer pairs were designed to target the V4 region of bacterial and archaeal 16S rRNA genes: forward primer 515F (GTGCCAGCMGCCGCGGTAA) with 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), and forward primer link (GT); reverse primer 806R (GGACTACHVGGGTWTCTAAT), with adapter (CAAGCAGAAGACGGCATACGAGAT), link (CC), and a unique 12-base specific barcode for each sequence. The PCR program included one cycle of 94 °C for 3 minutes, followed by 35 cycles at 94 °C for 45 seconds (denaturation), 55 °C for 1 minute (annealing), and 72 °C for 1 minute and a half (elongation), a final extension at 72 °C for 10 minutes, and then stored at 4 °C. After the program finished, the samples were run on an Agilent 2100 Bioanalyzer Instruments with Agilent DNA 7500 chips to ensure the amplicon quality and measure the DNA concentrations. Post PCR, amplicons were pooled (a set of 9 or 10) based on the peak height from Agilent DNA 7500 analyses and the concentrations were measured using KAPA Illumina kit for quantifying library (KK4824) from KAPA Biosystems. Samples were pooled again according to the results and number of sequences desired into a final one and the concentration was measured again with the KAPA quantifying kit. 4 nM pooled libraries were diluted with 10 mM Tris/0.05% tween buffer (pH = 8.5) and denatured with 0.2 N freshly made NaOH solution. The sample was then combined with 20% PhiX control kit to increase the library diversity. The mixture was incubated at 96 °C for 2 minutes using a heat block and then put into a water-ice bath for 5 minutes. After the sample was loaded, an Illumina MiSeq System with a MiSeq Reagent Kit v2 (300-cycles) and Metagenomic workflow was selected to execute the 16S protocol using the MiSeq Reporter software (MSR) at the University of Tennessee Genomics Core (UTGC).

2.5 Data analysis

Raw reads were quality filtered by QIIME2. Briefly, pair-ended sequences were joined and denoised by DADA2. A mapping of feature sequences was summarized, and a pre-trained Naive Bayes classifier based on Greengene 13_8 99% OTUs were used in the taxonomy analysis. A total of 196,7281 high-quality sequences were obtained from 16 samples. All samples were sub-sampled randomly at the minimal reads depth among all samples under `rarefy_even_depth` function in *phyloseq* package in R.

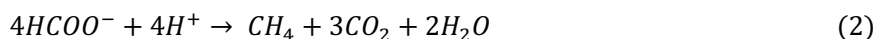
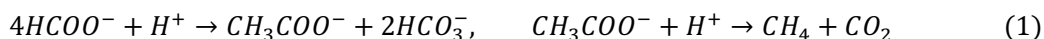
All the statistical analysis was conducted in R (Team 2016). Differences in the microbial community in different treatment stages were compared using analysis of variance (ANOVA) followed by Tukey's honestly significant difference (HSD) test. Pearson correlation analysis was finished under `corrplot` package in R (Racine 2012). K-means clustering was performed under `basic R`. Other plots were all performed under 'ggplot2' package (Ginestet 2011). The phylogenetic tree was plotted in `GraPhlAn` (Asnicar et al. 2015).

3. Results

3.1 Performance of the continuous reactors during formate spiking

For investigating the active microbial community of the anaerobic digestion under formate treatment, continued feeding of formate was carried out under no process perturbation, with a two-stage anaerobic digester operation during the study period (Figure 3-1a). Each stage started with the two-day loading of 10mM/day formate and followed by 5-day loading of 20mM/day formate.

In response to formate spiking, the changing trend of methane production was consistent with formate spiking level (Figure 3-1a). Theoretically, reactions involved in the conversion of formate to methane were provided in equations (1) or (2), 10mM formate would produce 63.6mL/L methane, while 20mM formate would produce 127.2mL/L methane. Consistent with the theoretical methane production, methane production increased gradually from 217±11 mL/L/day to 374±20 mL/L/day following the increase in formate from 10mM/day to 20mM/day. When formate loading was kept at 20mM, the methane production stayed stable around 360±12 mL/L/day. After the treatment stopped, the methane production decreased dramatically to 220±30 mL/L/day, the similar level as the control digesters. When the second spiking was performed, the methane production trend was similar to the first spiking. The methane production was following the logical volume. The actual methane production little differed from the theoretical volume indicated that the added formate had been completely utilized by the microorganism, then methane was produced.



3.2 Established *Methanosaeta* highly enriched in the formate spiked continuous reactors

Since methane production volume was as expected, functional methanogens were considered to have increased at first. The changing trend of methanogens in the digesters was shown in the barplot (Figure 3-1b). It is interesting to find that established *Methanosaeta* was the only archaea genus that changed accompanied by formate spiking (Figure 3-3), which is surprisingly different to the theoretical hypothesis. At the starting operation stage (stage1), *Methanosaeta* had its established abundance of 1.22%, then it increased to 3.88%, sustained above 2.44% during the formate spiking period, finally with a markedly drop after seven days recovery. Tukey-HSD test further identified the significant difference ($p<0.05$) of the relative abundance of *Methanosaeta* between spiking period (stage2-stage4) and none spiking period (stage1 and stage5). It is well known that *Methanosaeta* is popularly functioned during the acetoclastic methanogenesis process, and it is an acetate utilization specialist (Smith and Ingram-Smith 2007). Consequently, accompanied by *Methanosaeta* increased, the only acetate was found in the daily effluent, as suggested in Figure 3-1a. The changing trend of acetate was inconsistent with formate spiking and methane production. These findings of the solely significant shifts of *Methanosaeta* and effluent found acetate suggesting that formate has been transformed into acetate in the anaerobic process.

In details, one specific *Methanosaeta* (labeled as *Methanosaeta*#1) contributed most among all types of *Methanosaeta*, accounted for an average of 80% of *Methanosaeta* among the treatment period. *Methanosaeta*#3 also increased a little under treatment compared with the starting period, but always accounted for 16% of *Methanosaeta*. However, no increasing trend was found in *Methanosaeta*#2. The rare significance of *Methanosaeta*#2 and *Methanosaeta*#3 may suggest the critical function of *Methanosaeta*#1 in the digesters during the formate loading. Additionally, it was found that *Methanosaeta*#2 and *Methanosaeta*#3 showed a closer phylogenetic distance relation with *Methanosaeta*#1 (Figure 3-3).

Moreover, methanogens belonging to acetate utilizing methanogenic group (*Methanosacrinales*) were not absent in the digesters, *Methanosacrina* existed in the digesters with extremely low relative abundance, 0.02%~0.05%. *Methanosarcina* was suggested as more effective in using daily loading acetate and promoted more stable digestion than *Methanosaeta*, as suggested by (Conklin et al. 2006). It was also dominant in the mesophilic anaerobic digesters operated in the dairy farm (St-Pierre and Wright 2013). The stable low existence of

Methanosacrina may imply its non-function during the treatment, and the low competitiveness compared with *Methanosaeta*.

Hydrogenotrophic methanogens, belonged to two formate utilizing methanogenic orders (*Methanobacteriales* and *Methanomicrobiales*), were precisely presented in the digesters. However, the abundance of *Methanobacterium* and *Methanobrevibacter* (belongs to *Methanobacteriales*) were at low levels, with relative abundance smaller than 0.5%. The levels of *Methanospirillum* (belongs to *Methanomicrobiales*) remained around 0.5% throughout the operational period. The consistent low dominant and variant of formate utilizing methanogens may suggest the rare function of hydrogenotrophic methanogenesis, which may also imply formate was not directly transformed into CH₄ and CO₂.

3.3 Possible homoacetogens cooperated with established *Methanosaeta*

A total of 2333 unique representative reads spanning the V4 hypervariable region of the bacteria, and archaeal 16S rRNA gene was recovered from the three formate spiked continuous anaerobic livestock wastewater digesters. *Firmicutes* was the dominant phylum among 15 found phylum, constituted more than 30% of all detected reads (Figure 3-4), and it was mainly in part for family *Peptostreptococcaceae*, *Clostridiaceae*, and *Turicibacteraceae*. *Synergistetes* was the secondly ranked phylum, accounted for 13%~16% of all population, which was all from order *Synergistales*, family *Thermovirgaceae*, *Dethiosulfovibrionaceae*, and *Synergistaceae*. *Bacteroidetes* was the third-ranked phylum, with a proportion of 12%~16%, which was a small amount evenly distributed by population from order *Bacteroidales*. *Verrucomicrobia* and *Proteobacteria* were the subdominant groups, each comprising between 3% and 11% of the detections. These five bacterial groups represented approximately 73% ~ 80% of bacteria detected within the five samples.

Representative bacteria sequences with relative abundance greater than 0.1% and existed in at least 3/15 samples were chosen in the Pearson correlation analysis with *Methanosaeta*#1 and *Methanosaeta*. Bacteria with a significant *p*-value smaller than 0.05 remained in the corplot (Figure 3-5). Results showed that *f_Clostridiaceae*#8204 (representative sequence #8204 from family *Clostridiaceae*), *o_TSBW08*#2617 (representative sequence #2617 from order *TSBW0*), *c_noFP_H4*#11418 (representative sequence #11418 from class *noFP_H4*), *o_SHA-20*#9649 (representative sequence #9649 from family *SHA-20*) were all positive correlated with *Methanosaeta* and *Methanosaeta*#1, with correlation value all greater than 0.5, *p*-value smaller than 0.01. It indicated that those bacteria fluctuated on a similar trend as *Methanosaeta* and *Methanosaeta*#1. Since *Methanosaeta* is the possible methanogen that metabolizes formate, correlated bacteria may imply their possible corporations with *Methanosaeta* during the process.

3.4 Legacy effects redirected the pathway selection

For elucidating diversity relationship among bacteria, k-means clustering was applied in the later analysis. By k running from 1 to 10 to the relative abundance dataset, 4 clusters were best suggested by the minimal total within-sum of squares (Figure S3-1). With k assigned at 4, the reduction in the total within sum of squares would decrease to near 0. Therefore, we choose k=4 as the microbial population clustering, some bacteria scattered in cluster 1, cluster 2 and cluster 3, while the most rests were grouped in cluster 4 (Figure 3-6a). The relative abundance in the first 3 clusters was much higher than cluster 4. The four clusters were gradually changing with the population abundance. It may indicate that the population within different clusters may have different ecological behaviors.

Highly ranked communities with relative low-rank variance indicate the presence of legacy effects. Results showed that higher ranked taxonomic population could be consistently ranked high (Figure 3-6b), while the low abundance population was always at a low level in the digesters since raw data was used in the clustering, Figure 3-6b was then plotted based on the corresponding standard deviation with its mean rank which was grouped based on the previous clustering results. The first 2 clusters contain the top-ranked population, and also with relative stable rank variations (Figure 3-6c), a goodness of fit with $R^2=0.86$ suggested the higher ranked bacteria were more stable even a small number of the population are included. This does not mean top-ranked bacteria did not change, its relative abundance changed, but its abundance rank stayed top. That is to say, their rank

change was kept within a relative interval, the best was always the best. *Methanosaeta*#1 would be an excellent case to explain, as shaped in squared point in Figure 3-6b, it was at low rank and relatively low standard deviation. Combined with previous results, *Methanosaeta*#1 outcompeted any hydrogenotrophic methanogens and was responsive to formate addition. They always stayed within the rank of top 11. On the contrary, lower ranked population always at a low rank, and their distribution likely showed a pattern of stochastic. Since their changings were of relatively high-rank variance, which showed relatively higher rank standard deviation. No matter what this random fluctuation changes, they will not exceed a border and jumped into the array of the highly ranked population. As a consequence, no hydrogenotrophic methanogens excelled out during the treatment.

Pathway redirection is not just the status changing between *Methanosaeta* and other hydrogenotrophic methanogens, homoacetogens should be the ideal bridge between them. Besides methanogens, a few bacteria stand out among thousands of them. Former found possible homoacetogens all presented in the first three clusters. Especially, *f_Clostridiaceae*#8204 (representative sequence #8204 from family *Clostridiaceae*), a top-ranked population found in the digesters (Figure 3-6b), was established at high rank during the historical stage and also sustained at a very low variance. Its situation was similar to *Methanosaeta*#1. However, it increased with formate adding which shows a relative clear linear pattern (Figure 3-7). Likewise, *o_SHA-20#9649* and *c_TSBW08#2617* also showed similar trends (Figure 3-7). They were all clustered in the first 3 clusters with relative low rank and rank variance. In relative high abundance in the pre-spiking period, their relative abundance continued to increase with formate spiking. They may also take the same advantageous of their colonized priority as *Methanosaeta*, then helped *Methanosaeta* transform formate into acetate. Their non-coincidence existences may identify the existence of homoacetogenesis process in the digesters, the presence of homoacetogens may help understand redirection of the hydrogenotrophic pathway to the acetoclastic pathway. It would be those bacteria help to transform formate to acetate, and then the acetoclastic process happened, methane was produced then.

Further evidence of newly setup batch reactors further demonstrated the legacy enhancement in the continuous reactors. As described, the same substrate, formate, was also used as the only substrate to enrich hydrogenotrophic methanogens from anaerobic digesters treating livestock wastewater waste. The livestock water was from continuous reactors, which was in the same composition at the starting period. However, the only difference is the history of the reactor, continuous reactors have more than nine years of history, while the batch reactors have everything new. As a result, archaea accounted for 84.6%, while 84.1% of them were from family *Methanobacteriaceae*. This revealed a highly enriched proportion of the methanogens. It was not surprising to find that *Methanobacterium* was the most abundant genus in the reactor, represented by two unique feature sequences (respectively 72.5% and 5.0%), totaled up to 77.5% of the total population. The second abundant population was *Methanobrevibacter*, which accounted for 6.3% of the total reads (Figure 3-8). Both *Methanobacterium* and *Methanobrevibacter* were known to utilize formate as carbon source, which further indicated the primary hydrogenotrophic function of *Methanobacterium* under formate treatment. However, none sequences of those cooperated bacteria found in the continuous reactors were found in the newly setup reactors. It may further support the non-coincidence existence of legacy effects in the continuous reactors.

4. Discussion and Conclusion

Our study on how the microbial community responded to the formate, an essential but least discussed substrate in the anaerobic digestion system, which reveals that the importance of historical contingency on the contemporary microbial community structuring.

Formate has been recognized as a primary electron donor which can act independently of H_2 during the methanogenesis procedure. Formate can also be the interchangeable electron donors of H_2 during the catabolic process of methanogenesis. Similar to H_2 , it can be used by some hydrogenotrophic methanogens, even some hydrogenotrophic methanogens can use formate without the addition of H_2 . Therefore, discussion on what methanogens can make use of formate and what reasons influence the methanogenesis pathway selection is meaningful. We always speculate that formate would be prior utilized by hydrogenotrophic procedure, which not simultaneously happened with the acetoclastic process. Therefore, other pathway selection and functional community could have been neglected during the research. In this study, formate was utilized by methanogens, and methane was produced as expected. Surprisingly, the functional methanogen was not hydrogenotrophic methanogen but acetoclastic methanogen. *Methanosaeta*, especially *Methanosaeta*#1, was found in the active population. This result was different from the experiment under the same treatment but in different digesters, while hydrogenotrophic methanogens were the most responsive, and little overlap of functional methanogens was found between the two treatments (Figure 3-9b). In addition to methanogens, acetate, not formate, was the main volatile fatty acids in the daily effluent from the triplicate digesters. It was a further indication of formate to acetate transformation. Furthermore, our previous experimental results showed that the established *Methanosaeta*#1 found under formate spiking were mainly in the same sequences with acetate spiking in the same reactors (Figure 3-9a), which mainly played the acetoclastic methanogenesis function.

In addition, the connection bridge between the hydrogenotrophic pathway and the acetoclastic pathway was a group of bacteria, homoacetogenic bacteria. These bacteria were characterized as chemotropic, H_2 and CO_2 users, had high thermodynamic efficiencies in its metabolism. It could grow on one-carbon and multi-carbon compounds (Novaes 1986). Homoacetogens, such as *Moorella thermoacetica*, *Moorella thermoautotrophica*, could be functioned as very versatile anaerobes that convert a variety of different substrates to acetate as the primary end product (Diekert and Wohlfarth 1994). *Clostridium ultunense* was mentioned as the homoacetogenic bacteria, which grow autotrophically on H_2/CO_2 or carbon monoxide (CO) and/or heterotrophically on formate as sources of energy and carbon (Schink 1997, Schnürer et al. 1996, Xiao et al. 2017). (Gonzalez-Fernandez et al. 2015, Xiao et al. 2017) and (Schink 1997) also showed that *Clostridium*, a possible genus, would transform formate into acetate. They all showed the highest possible existence of formate-acetate conversion in the anaerobic environment. However, it is interesting that some possible homoacetogenic bacteria existed, a good case is a population from family *Clostridiaceae*. Family *Clostridiaceae* contains the *Clostridium* genus, which has been identified as the critical homoacetogenic bacteria (Molitor et al. 2016). The enhancement of homoacetogenesis can result in acetate increase happened under the H_2 pressure. Microbial populations and operational conditions were the major factors that influence the metabolic pathways. However, our experiment showed the bacterial community at the starting time, and the operational conditions were all the same, how could that happen in our digesters, could that be any other factors we still have not paid attention to?

As a consequence, the highly ranked population were most responsive but sustained at the top rank, which could be an indication of the influence of past community structure, here we preclude as legacy effects. “Legacy effects,” an ecology terminology, which was first applied to the plant study in the 1990s (Amaranthus and Perry 1989, Molina and Amaranthus 1990, Perry et al. 1987), was more focused on rhizosphere “legacy.” Its core value was that the activity of the past potentially influences the current microbial ecology. It could also be explained that the current situation is a snapshot of one specific lifetime for a specific environment, and it is significantly imprinted by the abiotic or biotic conditions shaped during the pre-lifetime. Previous literature also summarized the mechanisms of the legacy effects were time-lags in species sorting and priority effects. Two mechanisms would result in four scenarios, no legacy effects, transient legacy effects, persistent legacy effects, and mixed scenario (Vass and Langenheder 2017). It has been widely studied in the field of soil, plant, forest, vegetation, land in the ecosystem. How can past ecology make optimal use of the available sources? To take advantage of the newly coming substrate, *Methanosaeta*, the “home” archaea in the digesters, cooperated with “home” homoacetogens, then transformed the contemporary sources into acetate, their preferable substrate. This

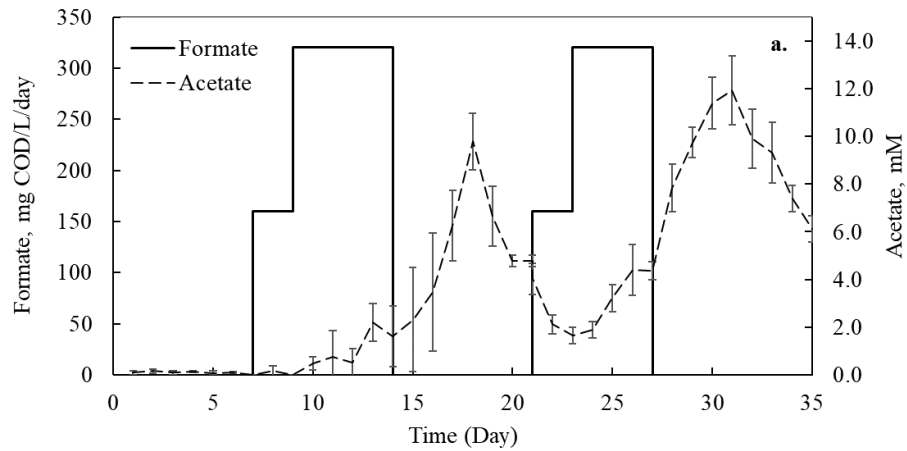
deeply revealed the vast extent of the pre-optimized microbial community has the priority in controlling the anaerobic digestion process. The recognized acetoclastic methanogenesis experts, *Methanosaeta*, was significantly increased together with the available substrate. However, the widely known formate utilization methanogens, hydrogenotrophic methanogens, did not increase. It could be that *Methanosaeta* was the early principal settler in the digesters, which has already formed the so-called monopolization power in resisting any other microbial community making use of the coming substrate and growing. By contrast, nothing related to the legacy effects was found in the newly setup batch reactors, even they had the same initial microbial community and the same substrate. It may be because the new environment was the same for most microbial population, nothing heritages functions were there for them. Therefore, the theoretically functional community, hydrogenotrophic methanogens, were enhanced and reinforced.

Moreover, the legacy effects of homoacetogen was also found. On the one hand, “home” bacteria *f_Clostridiaceae#8204* (representative sequence belongs to phylum *Firmicutes*, class *Clostridia*, family *Clostridiaceae*) sustained at the high rank and low-rank variance. On the other hand, “home” bacteria and “home” methanogen coexisted and simultaneously worked together, which made the pathway redirected.

In general, substrates, environmental parameters, and some other factors have been emphasized as the essential parameters, while in our experiment even those factors were the same, the results were entirely different. Much attention has been paid to snapshot experiment itself, but sometimes the vital factor, legacy effects, has been neglected. Therefore, in the research field of anaerobic digestion, how historical community structure would impact the current experiment remains to be discussed more in the future.

Appendix C: Figures and Tables

(a)



(b)

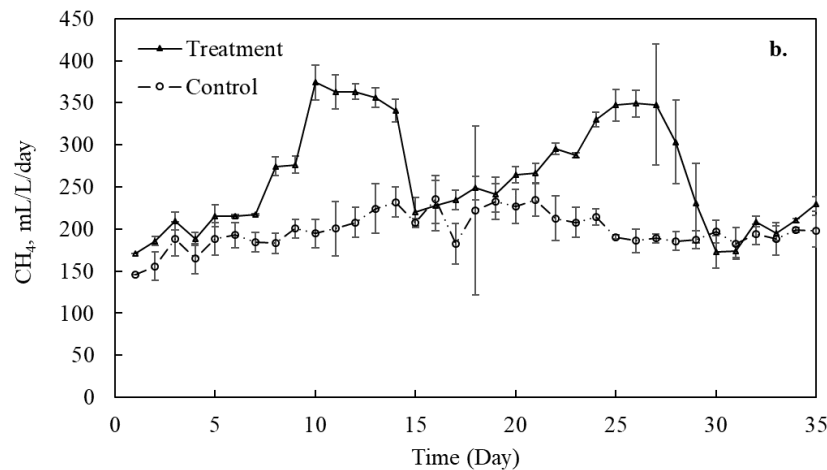


Figure 3-1 Figures of performance parameters.

(a) Formate spiking concentration and effluent acetate concentration in the daily effluent from the digesters. Cross label with the straight line showing formate concentration in the spiking. Circle label with the dashed line indicating the concentration of acetate in the effluents. (b) Methane production during anaerobic treatment process with formate spiking. Results are means of triplicates with the error bars representing standard deviations. Triangle label with the straight line indicating the CH₄ production level under formate spiking. Circle label with the dashed line showing the CH₄ production level under non-treatment.

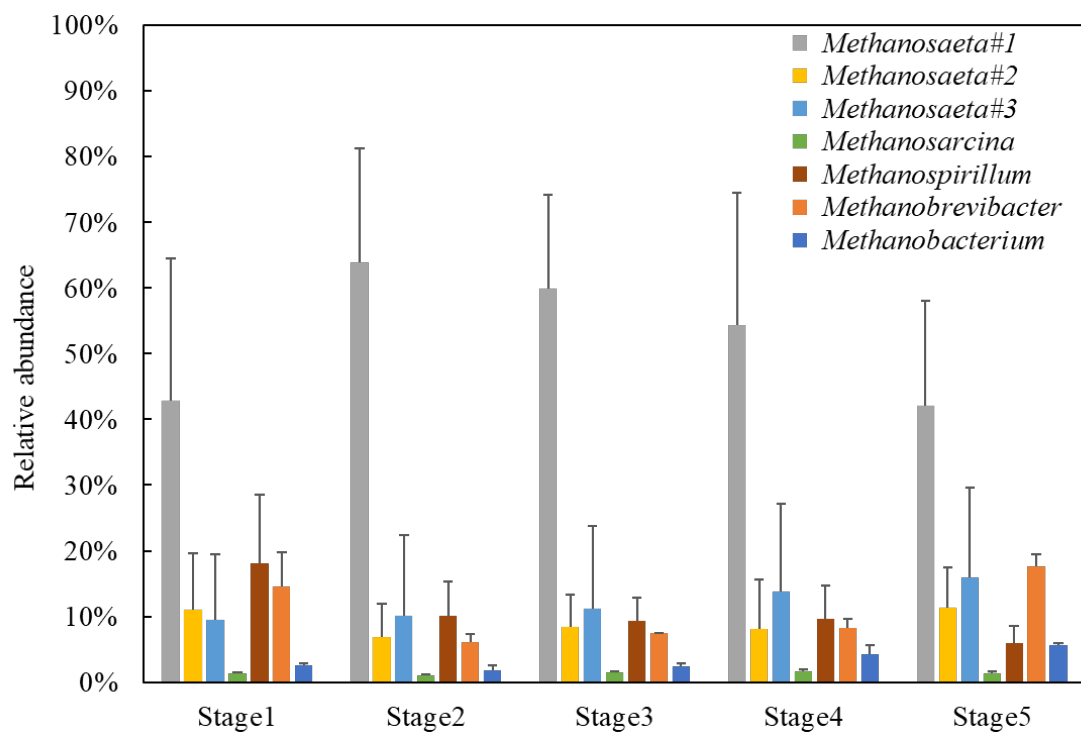


Figure 3-2 Barplot of methanogen relative abundance profile during the formate spiking.

Data are means of triplicates samples. *Methanosaeta*#1 means #1 sequence variant is classified as *Methanosaeta*. The same applies to *Methanosaeta*#2 and *Methanosaeta*#3. Error bars are standard deviations.

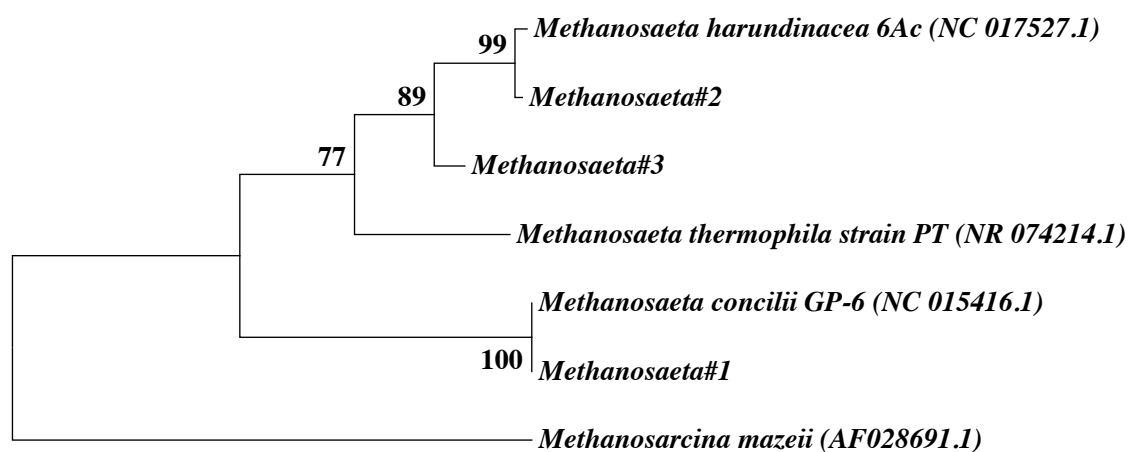


Figure 3-3 The relationships between full-length archaeal 16S rRNA gene sequences and different *Methanosaeta* in the digesters explained by neighbor-joining phylogenetic tree.

The tree was plotted in MEGA7. Bootstrap value is shown next to the branches. The evolutionary distances were estimated with the Maximum Composite Likelihood method. The scale bar represented the number of substitutions per sequence position. The number indicated in the parentheses next to the sequence designations are GenBank accession numbers of the 16S rRNA gene sequences.

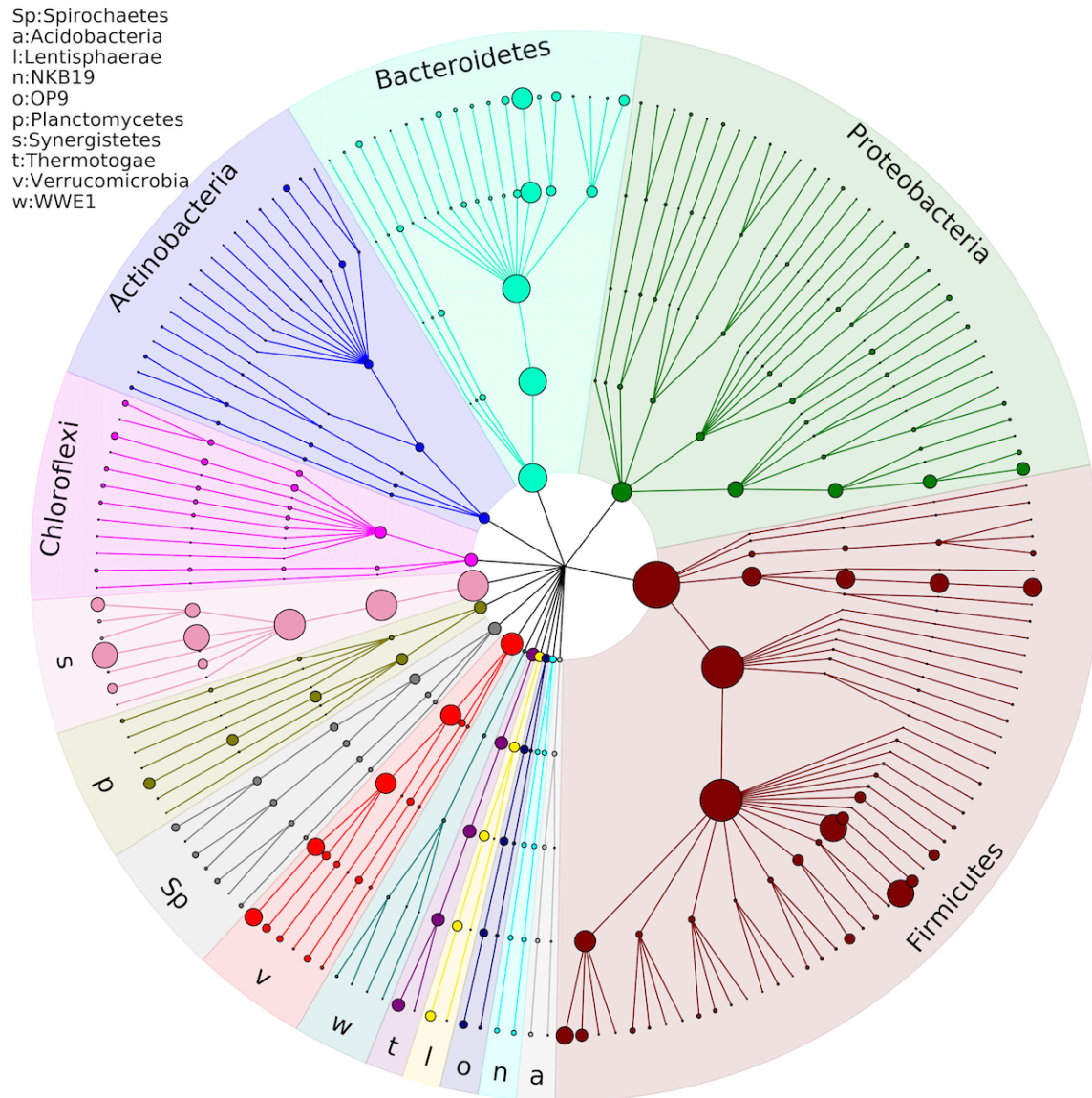


Figure 3-4 Phylogenetic tree describes the bacterial composition in the digesters.

Results are means of triplicate digesters. It depicts the distribution from phylum to genus. Different branches and annotation background stand for different phylum. The more considerable annotation means the more abundant population. The legend describes the phylum name at the end of the clades. The plot is constructed in MetaPhlAn2.

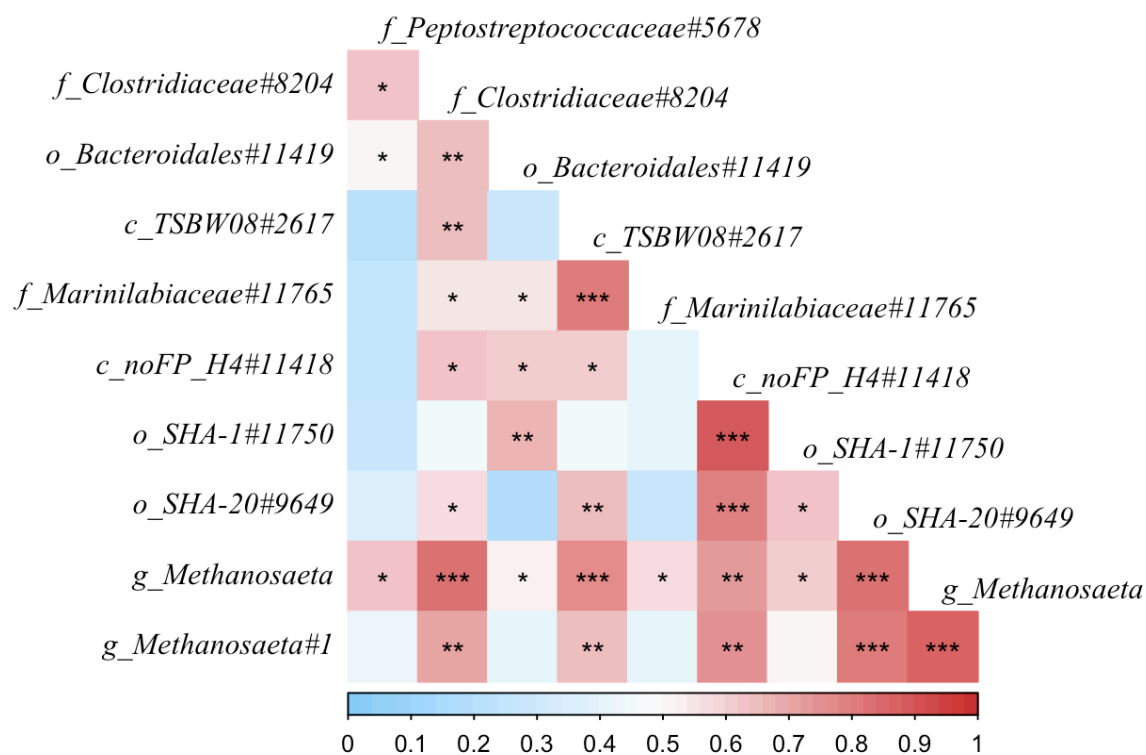


Figure 3-5 *Methanosaeta* correlated bacteria found during the formate treatment in the continuous digesters.

The stars ***, **, * respectively stands for p -value at significant level 0.001, 0.01, 0.05. The filling color stands for the correlation value (r), as shown in the scale bar.

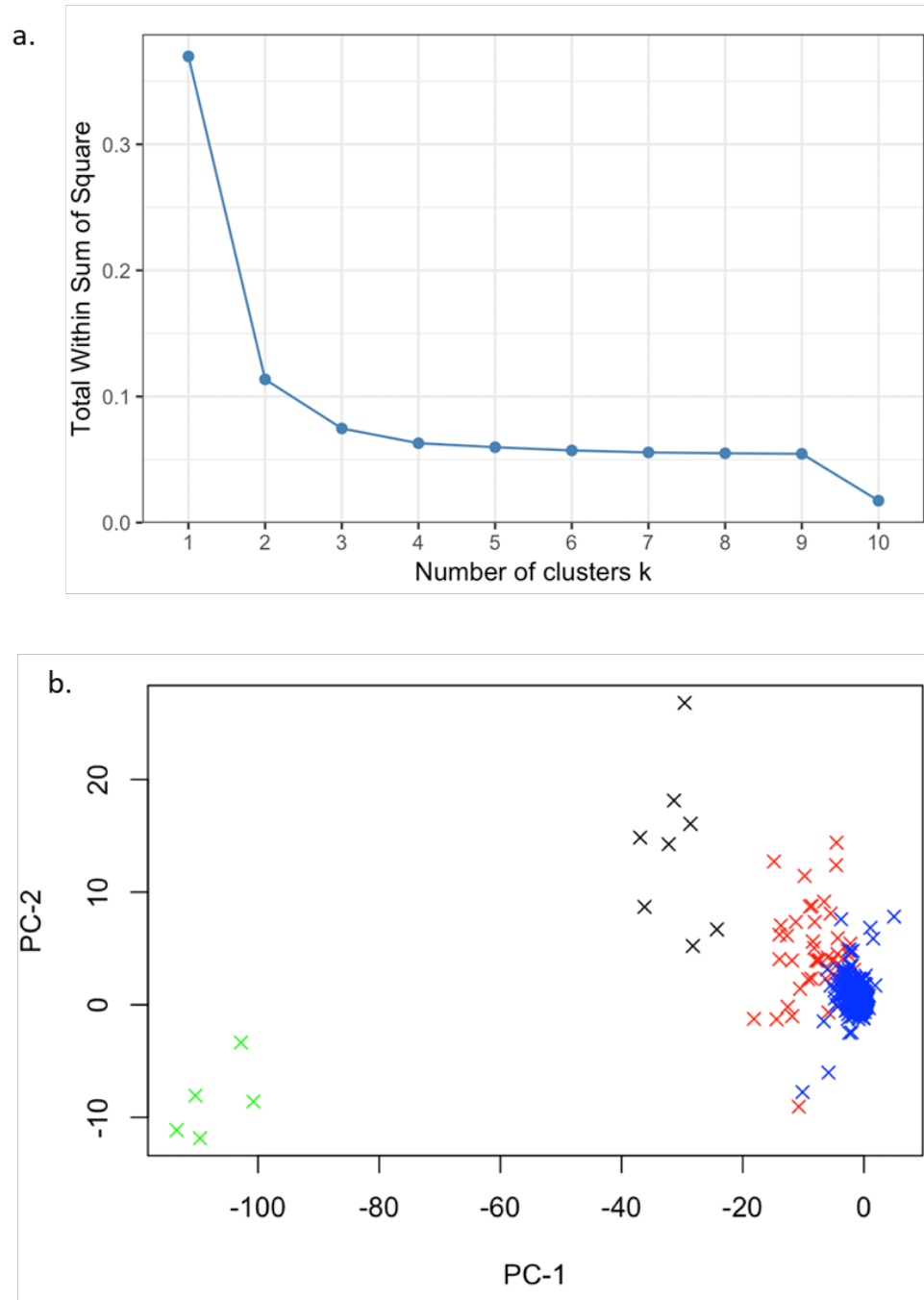


Figure S3-1 K-means clustering of the microbial abundance results.

a. Determining the optimal number of clusters for k-means clustering. Scree plot showing the elbow method (total within sum of squares) in best k determination. b. Visualization of the clustering result.

Figure 3-6 Result of k-means clustering analysis of the taxonomic population.

a. The distribution of the microbial abundance of the four produced clusters. X-Axis showing the relative abundance of the microbial community at sequence variant level. Y-Axis displaying the kernel density estimation. In this plot, the Gaussian kernel is drawn at every individual data point, and all of these curves are then added together to make a single smooth Gaussian density estimation. Since cluster 4 contains too many data points with an abundance near 0, the microbial community with relative abundance smaller than 0.0005 are not displayed in the plot for a better distribution visualization. The legend in this plot applies to b, c, and d. **b.** is the rank-based scatter plot, which shows the relation between the mean rank of microbial abundance and the standard deviation of microbial rank during the study period. Rank stands for the degree of abundance of each sequence variant in the sample. For example, if one sample contains 2000 sequences variants, rank=1 means the most abundant microbial community in the digester, rank=2000 represents the least abundant microbial community. X-Axis showing each sequence variant's rank in the dataset, y-axis stands for the variation of each bacteria's rank change. **c.** The scatter plot of the first three clusters. It shows the relationship between rank and the standard deviation of rank. $r=0.86$ with $p<0.001$ is the result of Pearson correlation analysis of the scattered points. The fitting curve is the linear fitting of the scattered points. The grey area is the standard error. **d.** is the zoomed-in plot of the high ranked and low variational microbial community. All of them belongs to cluster1, cluster2 or cluster3. Black labeled points are the microbial community with significant correlation with *Methanosaeta*.

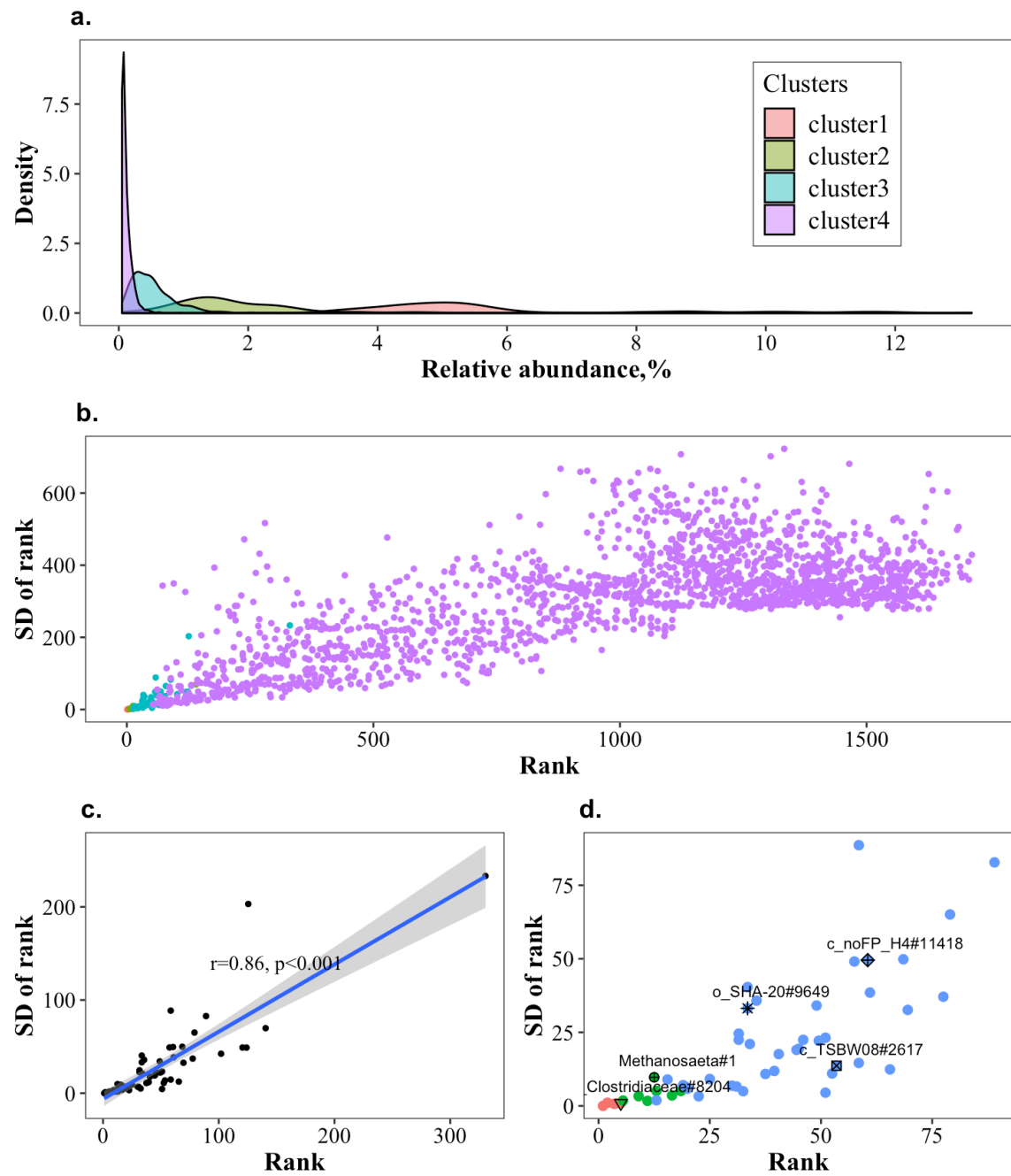


Table 3-1 Microbial community clustered into cluster 1, cluster 2 and cluster 3.

cluster	microbial community	cluster	microbial community
cluster1	f_Thermovirgaceae#5997	cluster3	g_Syntrophomonas#8659
cluster1	g_Turicibacter#384	cluster3	f_Peptostreptococcaceae#248
cluster1	f_Peptostreptococcaceae#5678	cluster3	o_Bacteroidales#10870
cluster1	f_Peptostreptococcaceae#6504	cluster3	o_Bacteroidales#10855
cluster1	f_Clostridiaceae#8204	cluster3	f_Marinilabiaceae#11926
cluster2	o_WCHB1-41#2931	cluster3	o_WCHB1-41#9737
cluster2	g_Syntrophus#11271	cluster3	g_Treponema#1892
cluster2	g_Methanosaeta#1485	cluster3	g_Acholeplasma#10603
cluster2	g_Kosmotoga#3206	cluster3	o_Bacteroidales#11419
cluster2	f_Bacteroidaceae#642	cluster3	o_WCHB1-41#10516
cluster2	f_Pirellulaceae#4890	cluster3	f_Pseudomonadaceae#10146
cluster2	f_Peptostreptococcaceae#4754	cluster3	f_RFP12#5167
cluster2	o_Bacteroidales#3664	cluster3	f_Marinilabiaceae#11765
cluster3	g_HA73#8178	cluster3	g_Sedimentibacter#10598
cluster3	o_SHA-1#11750	cluster3	c_TSBW08#2617
cluster3	g_Kosmotoga#30	cluster3	c_noFP_H4#11418
cluster3	g_vadinCA02#6498	cluster3	g_Sphaerochaeta#4596
cluster3	f_R4-41B#260	cluster3	o_Bacteroidales#5688
cluster3	o_SHA-20#9649	cluster3	g_Syntrophus#8840
cluster3	g_Syntrophomonas#9019	cluster3	g_Methanosaeta#7182
cluster3	g_Clostridium#7567	cluster3	o_Clostridiales#6193
cluster3	g_HA73#8425	cluster3	g_Methanospirillum#5213
cluster3	g_Syntrophomonas#8713	cluster3	o_Bacteroidales#4907
cluster3	f_R4-45B#9502	cluster3	g_Halomonas#684
cluster3	g_HA73#12248	cluster3	o_WCHB1-41#497
cluster3	f_Ruminococcaceae#6928	cluster3	f_RFP12#4190
cluster3	o_Bacteroidales#3188	cluster3	g_Corynebacterium#366
cluster3	o_Bacteroidales#11408	cluster3	g_HA73#3472
cluster3	o_WCHB1-41#11187		

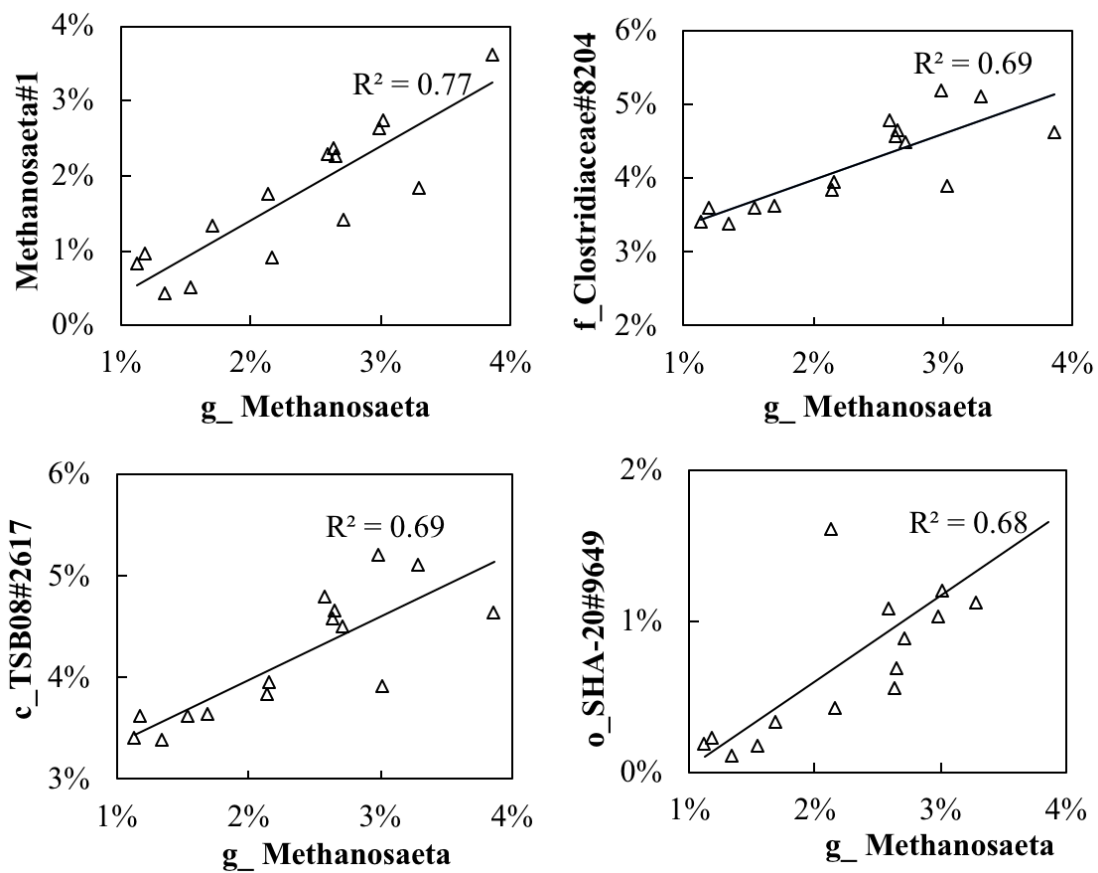


Figure 3-7 The changing trend of the possible homoacetogenic bacteria with *Methanosaeta*.

Data are means of triplicate digesters.

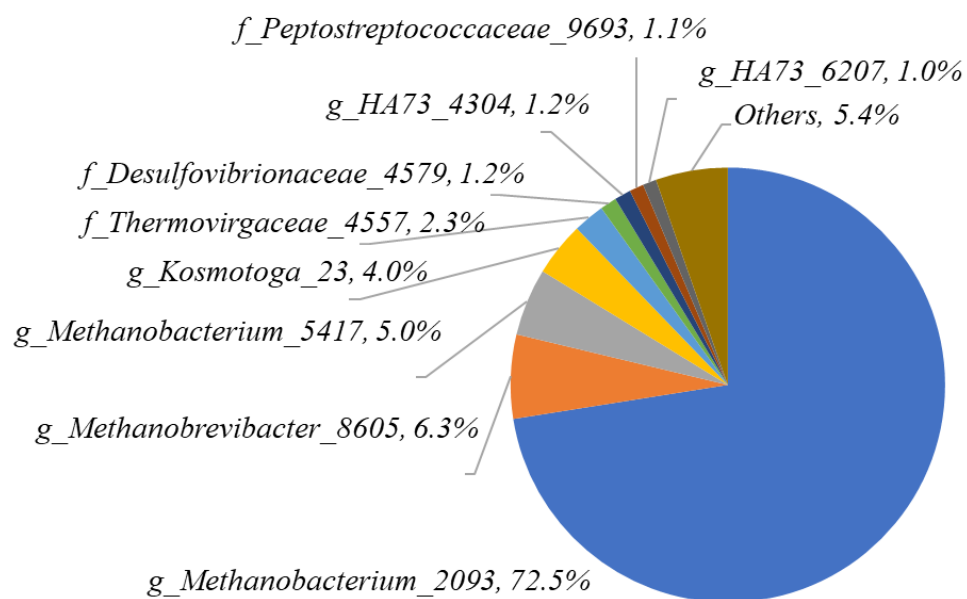
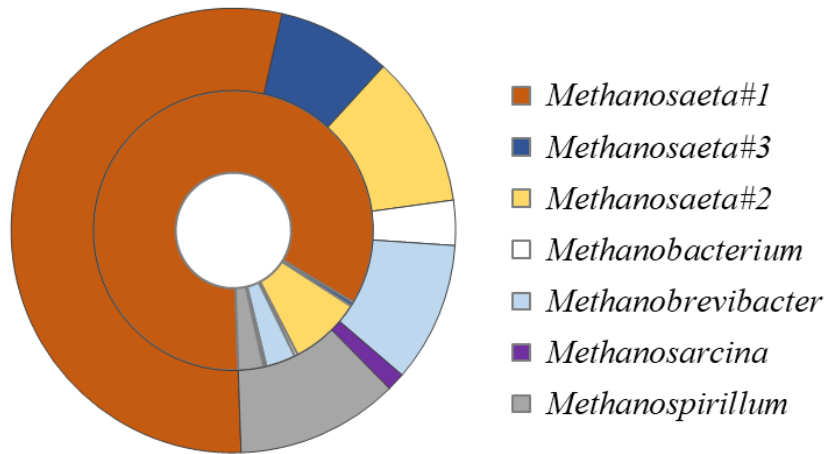


Figure 3-8 Phylogenetic distribution of the microbial community in newly formate batch reactors.

“*g_Methanobacterium_2093*, 72.5%” stand for the representative sequence from genus *Methanobacterium* and the representative sequence number is 2093, and the relative abundance is 72.5%. The same applies to the other labels.

a. Different treatment in the same reactors



b. Different reactors with the same treatment

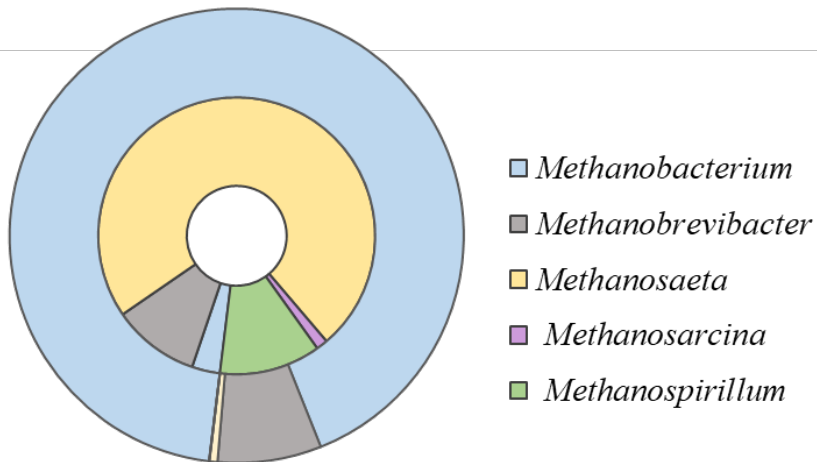


Figure 3-9 Methanogen composition comparison between acetate and formate treatment.

Data are means of triplicate digesters. a. compares the archaea composition between formate and acetate spiking in the continuous reactors. “A” stands for acetate treatment, while “F” stands for formate treatment. b compares the archaea composition under formate treatment in different digesters. “C” stands for formate treatment in the continuous reactors, while “B” stands for formate treatment in batch reactors.

Section B: Anaerobic Membrane Bioreactor Treatment for the Treatment of Recalcitrant Biofuel Process Wastewater

Abstract

Pyrolysis of lignocellulosic biomass is one of the options to produce biofuel as a renewable source of energy. The process wastewater from the catalytic pyrolysis of biomass contains a significant amount of soluble organic carbon, with the COD of such wastewater frequently exceeding 15,000 mg/L. The high organic content of the process wastewater presents an opportunity to recover energy using technologies such as anaerobic digestion. However, microbial activities underlying the anaerobic digestion processes could be hindered by the toxicity of certain wastewater constituents, including heterocycles, phenolics, and cycloalkane, as well as the high acidity of this type of wastewater. A laboratory-scale anaerobic membrane bioreactor (AnMBR) was developed to enrich microbial biomass both resistant to the toxicity and capable of converting the organics into methane. With the gradual increase in organic loading of the process wastewater, the AnMBR was able to remove >99% of COD loading with stable biogas production, which coincided with the dominance of *Methanobacterium*, suggesting the functional importance of hydrogenotrophic methanogenesis in the anaerobic metabolism of the wastewater organics. Other methanogen populations, including *Methanosaeta* and *Methanomassiliicoccus*, were also found at significant levels in the AnMBR, indicative of the need for functional diversity to support the degradation of diverse organic constituents in the process wastewater. More importantly, the prevalence of multiple syntrophic bacteria (*Syntrophaceticus*, *Syntrophobacter*, *Syntrophomonas*, *Syntrophorhabdus*, and *Syntrophus*) demonstrates the importance of syntrophic interactions in anaerobic treatment.

Keywords: anaerobic membrane bioreactor; wastewater; lignocellulosic; methanogen; methane

1. Introduction

Immense opportunities, such as the increasing awareness on low ecological footprint buildings, hydrogen-fueled cars, buses, and other transport machinery, have been offered on renewable energy (Nowotny et al. 2018). Biofuel is a vital source of renewable energy. In the first generation of biorefinery, biofuels dominantly refined from food stock, such as corn, vegetable oils. In the second generation, feedstocks shifted into lingo-cellulosic materials, such as residues from agriculture and forestry and energy crops, woodchips. Lignocellulose can also be renewable source of the biofuel production. It can help overcome the competition of increased food demand for the arable land, provide an alternative to petroleum dependent products, and contribute significantly to the global energy supply in the future. Thus, it contributes significantly to the energy security and allows a decreased influence on the environment (Jonsson et al. 2013).

Biomass pyrolysis is a promising approach, which can help deconstruct lignocellulosic substrates to intermediates for upgrading to biofuel or aromatic chemicals (Black et al. 2016). During the biomass pyrolysis process, however, nearly nine times of water addition than biofuel are required to partition the biofuel from the aqueous stream (Vispute and Huber 2009). Therefore, a large amount of wastewater at high COD concentration and low pH with considerable amounts of compounds will be generated (Thomsen et al. 2009), which attributed to the toxicity of the wastewater. The anaerobic digestion process is a beneficial carbon recovery process. It requires minimum energy and chemicals, low excess sludge production, and the final product is methane that can be widely used for energy or hydrogen production in the integrated biofuels process. It has been proved to be capable of dealing with a large amount of industrial wastewater treatment (Lew et al. 2009). However, the anaerobic treatment process does not suffer from a wide variety of inhibitory substances, while they are present in substantial concentrations in wastes (Chen et al. 2008). Therefore, the high toxicity of biofuel wastewater is likely challenging to treat anaerobically. The impact of the toxicant is related to the amount of the compound and the amount of biomass (sludge) inside the reactors. Moreover, the toxicity and inhibition are intrinsically related to the process conditions (Chen et al. 2008). The predominant condition in anaerobic digestion is the microbiome adaptation of biomass. The absence of specific microbial populations results in the unbalanced anaerobic food web, which leads to metabolic imbalance and process failure. Thus, balanced anaerobic food web requires the right microbes with the correct quantity for each target substrate and intermediate. Therefore, we need to rapidly establish a microbial consortium, which is capable of effectively degrading biomass to methane.

Anaerobic membrane bioreactor (AnMBR) can ensure complete solids retention and process stability, which is capable of handling high organic loadings and mixing intensities. The membrane in an AnMBR can offer advantages for slow-growing microorganisms with an increase in their retention ability (Svojitka et al. 2017). Poor retention of biomass in the reactor and high residual effluent pollution pose negatively effect on the industrial wastewater treatment (Lin et al. 2013). Our result showed that AnMBR has a strong capability in biomass retention and methane production. Thus, the present work aims to determine whether the operation strategies can help develop a group of microbial community to thrive and facilitate the biological conversion of biofuel wastes to methane. Therefore, the effort involved an investigation into what microbial populations can thrive with the increasing loading of toxicity biofuel process wastewater.

2. Materials and Methods

2.1 Wastewater pre-treatment

Since the biofuel wastewater is toxic to the anaerobic digestion process, biodegradable organics was used as the leading organic feed, and the biofuel wastewater was added to be co-digested. The biofuel wastewater was acidic (pH = 2.3), which was later neutralized with 10N NaOH before adding to the feed (800mL biofuel wastewater + 24mL 10N NaOH). For the primary carbon source, oil (Soybean oil), alcohol (Beer), protein (Skim Milk) and carbohydrate (Starch) were added in the feed. Yeast extract and other micronutrients (Met Source AN & Granule Aid, River Bend Labs) were added. Alkalinity was added using NaHCO₃ and KHCO₃. The amount of each component in the 1 L feed is presented in Table 4-1.

2.2 Anaerobic membrane bioreactor set-up

A laboratory scale AnMBR reactor was prepared in order to test the applicability of the biofuel wastewater for the methane production process. The schematic diagram of the AnMBR system is presented in Figure 4-1. A benchtop bioreactor from Eppendorf (Bioflo®&Celligen®310, Eppendorf) was used for the anaerobic reactor, and a cross flow tubular membrane fiber (X-flow, Pentair) was used as the membrane system to separate treated water from anaerobic sludge. Since the filtration type was a side stream and cross-flow type, no biogas scouring was applied. The seed sludge was provided from a biomethane system (Veolia).

The bioreactor tank volume was maintained at 9-10 L and the feed flow rate was set at 2mL/min. The anaerobic reactor was maintained at 35°C (mesophilic digestion). Oxidation-reduction potential, pH, and temperature were continuously monitored, and two-level switches were equipped and connected to the feed and recirculation pumps in order to prevent any overflow or depletion of the tank. Two peristaltic pumps were installed for feed supply (Masterflex® L/S® Standard Digital Drive) and membrane recirculation (Masterflex I/P Precision Brushless Drive with head, 77410-10). The sludge wasting frequency was about 450mL per week, and the total reaction volume was 10L. The SRT was maintained at 156 days.

One tubular membrane fiber was installed in the glass membrane housing and the sludge was pumped into the membrane at 2.55 L/min to maintain the shear velocity at 2m/s. The filtrate flow rate was maintained at 2g/min using the mass flow controller (mini CORI-Flow, Bronkhorst). At this filtrate flow rate, the membrane flux was calculated at 8.4 LMH. Three pressure transducers (IFM) were installed to monitor the inlet, outlet and filtrate pressure. Recirculation direction was reversed daily to mitigate the fouling on the membrane. CIP (Clean-in-place) was conducted weekly using NaOCl and Citric acid.

2.3 Chemical analysis

Daily biogas production of the anaerobic digesters the primary parameter was used to evaluate the anaerobic performance, which was determined by the mass flowmeter (Alicat mass flow meter, MS-100SCCM-D/5M). Biogas component was measured using the Micro GC (3000A, Agilent). Chemical Oxygen Demand (COD) was measured using the COD test kit (COD TNTplus, Hach).

2.4 DNA extraction and 16S rRNA amplicon sequencing

Biomass samples were collected from the digesters periodically based on the loading rate. All samples were stored at -80 °C until DNA extraction. DNA extractions were performed using MP FastDNA™ Spin Kit for Soil according to the manufacturer's instruction (Protocol Revision #116560200-201608). Briefly, the biomass samples were resuspended in the 978 µl Sodium Phosphate Buffer solution and 122 µl MT Buffer, then followed by 100 seconds vortex on MP biomedical fast prep instrument. Then the DNA was purified with Zymo Genomic DNA Clean & Concentrator™-10 kit according to the instructions provided by the manufacturer. The DNA concentrations were determined using Thermo Scientific NanoDrop ND-3300 Fluorospectrometer.

Polymerase chain reactions were prepared with a cocktail mix containing 12.5 µL Phusion flash Master Mix, 10 µL ultra-pure water, 1 µL forward primer, 1 µL reverse primer, and 2 µL (100 to 150 ng) DNA template. Each universal primer pairs were designed to target the V4 region of bacterial and archaeal 16S rRNA genes: forward primer 515F (GTGCCAGCMGCCGCGGTAA) with 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), and forward primer link (GT); reverse primer 806R (GGACTACHVGGGTWTCTAAT), with adapter (CAAGCAGAAGACGGCATACGAGAT), link (CC), and a unique 12-base specific barcode for each sequence. The PCR program included one cycle of 94 °C for 3 minutes, followed by 35 cycles at 94 °C for 45 seconds (denaturation), 55 °C for 1 minute (annealing), and 72 °C for 1 minute and a half (elongation), a final extension at 72 °C for 10 minutes, and then stored at 4 °C. After the program finished, the samples were run on an Agilent 2100 Bioanalyzer Instruments with Agilent DNA 7500 chips to ensure the amplicon quality and measure the DNA concentrations. Post PCR, amplicons were pooled (a set of 9 or 10) based on the peak height from Agilent DNA 7500 analyses and the concentrations were measured using KAPA Illumina kit for quantifying library (KK4824) from KAPA Biosystems. Samples were pooled again according to the results and number of sequences desired into a final one and the concentration was measured again with the KAPA quantifying kit. 4 nM pooled libraries were diluted with 10 mM Tris/0.05% tween buffer (pH = 8.5) and denatured with 0.2 N freshly made NaOH solution. The sample was then combined with 20% PhiX control kit to increase the library diversity. The mixture was incubated at 96 °C for 2 minutes using a heat block and then put into a water-ice bath for 5 minutes. After the sample was loaded, an Illumina MiSeq System with a MiSeq Reagent Kit v2 (300-cycles) and Metagenomic workflow was selected to execute the 16S protocol using the MiSeq Reporter software (MSR) at the University of Tennessee Genomics Core (UTGC).

2.5 Sequencing data analysis

The raw data were quality filtered by QIIME2. Briefly, pair-ended sequences were joined and denoised by DADA2. A mapping of feature sequences was summarized, and a pre-trained Naive Bayes classifier based on Silva 132 99% OTUs were used in the taxonomy analysis. Unique sequences classified from the DADA2 were labelled as sequence variants (SVs). All samples were sub-sampled randomly to the minimal read number among all samples by *rarefy_even_depth* function in phyloseq package in R.

3. Results

3.1 Test on compound composition for biofuel wastewater toxicity

High COD concentration of the biofuel wastewater offers potential opportunity for anaerobic treatment. However, anaerobic digester set-up requires the strict condition, such as pH, accumulation of organic acids, and etc., which could inhibit the biogas production and cause the unstable reactor performance. Therefore, we did a test to identify the biofuel wastewater compounds composition. As a result, the biofuel process wastewater was mainly detected with 295 μ g/mL of methanol, 386 mg/L formaldehyde, 237 mg/L acetaldehyde, 263 mg/L acetic acid and other organic compounds (Table 4-2). Such toxicity raised the problem of the anaerobic reactor set-up. Therefore, prior to the continuous operation of AnMBR treatment, the biofuel process wastewater was analyzed for toxicity by testing methane production in methanogenic cultures under various loadings of biofuel process wastewater.

Different loadings of biofuel process wastewater, ie: 0.01, 0.02, 0.05, 0.1, 0.2 and 0.5 %(vol/vol), were amended into test reactors (Figure 4-2a). The control and treatment batch reactors at each loading rate were set in duplicate and incubated into 35°C. The biogas production and methane concentration were monitored continuously along the experimental period. With the increase of the biofuel process wastewater loadings, its inhibitory effects on methane production significantly increased (Figure 4-2a). With the loading rate of biofuel process wastewater at 0.02% (vol/vol), methane production comparing with control reactors was reduced. When the biofuel process wastewater loading continuously got elevated, the methane production was totally ceased, which indicated that certain chemicals in biofuel process wastewater were severely toxic for microorganisms. Biofuel wastewater could cause serious inhibition in methanogenesis procedure with loading rates as low as 0.02% and could lead to a complete inhibition of methanogenesis with loading rate as 0.05% or above.

Moreover, the comparison of methane-producing rate (microorganism growth rate) of reactors with different biofuel process wastewater loadings during the exponential phase also demonstrated the inhibitory or toxic effects clearly (Figure 4-2b). The methane production rates between control and treatment reactors with 0.01% biofuel process wastewater are similar, 2.05 and 1.98mL, CH₄/hr, respectively. With the increase of biofuel process wastewater to 0.02%, however, the methane production rate decreases to 0.21mL/hr dramatically (Figure 4-2b), which indicates that threshold of biofuel process wastewater spiked into anaerobic digesters without inhibitory effects on microorganisms is about 0.01% (vol/vol). Other loading rates above this value may cause severe inhibitions. From the evaluation of the toxicity test of biofuel process wastewater on anaerobic digestion, it indicates that the loading rate of this waste applied into anaerobic digestion should be well designed and monitored.

3.2 Process adaptation to increased loading of toxicity biofuel process wastewater

Anaerobic membrane bioreactor (AnMBR) was chosen as the model system to control the adaptiveness by gradually increasing biofuel wastewater retention time. In this study, we set up a laboratory scale AnMBR system to study the biomass to methane conversion. Based on the preliminary toxicity test result, a biodegradable supplemental feed was designed to provide enough buffering ability (alkalinity), various types of biodegradable carbon source (alcohol, fat, protein and carbohydrate) and nutrients to a diverse of microbial populations. 10M NaOH was neutralized with the biofuel wastewater, and the portions of biofuel wastewater in the feed were gradually increased from 0.02% to 1.8% (vol/vol).

In the initial 2-month start-up study period, a low COD removal rate was found in the system. From T2, COD removal rate increased and gradual sustained at a stable state, suggesting the change of main microbial consortium. With 1-year continuous operation, we gradually increased the biofuel wastewater concentration, the portion of the biofuel wastewater was increased to 1.8% of total feed volume, which is 90 times the initial inhibitory level. Even though the portion of the biofuel wastewater was as low as 1.8% of the total volume, COD contributed to nearly 50% of the total COD feeding in the biofuel wastewater since the COD of biofuel wastewater was much higher than the COD of biodegradable wastewater. The COD in feeding of biofuel

wastewater was around 10,654 mg/L, and the COD of biofuel wastewater was around 5,000 mg/L. The COD concentration in the effluent was sustained at 1,200 mg/L after one year of slow acclimation. However, analysis of the AnMBR filtrate samples indicated that neither of these analytes present. If we assume that the COD of synthetic wastewater was completely consumed in the anaerobic process, it suggests that about 80% of biofuel wastewater was successfully converted to biogas. The daily COD removal rate was calculated using the daily COD feeding and permeate COD, and the results were compared with the actual daily COD loading rate (Figure 4-3), which revealed that more than 90% of COD are converted to methane gas.

The carbon source in the biofuel wastewater can be successfully recovered in the AnMBR system biologically. The designed AnMBR system improves the overall efficiency of the biofuel production. In this recovery procedure, the synthetic wastewater used in the mitigation of toxic biofuel wastewater together with supplied alkalinity, nutrients and biodegradable carbon source were essential conditions that can help convert biomass in the biofuel wastewater into methane. The adaptable microbial community is the core of the anaerobic biogas conversion process, which involves multiple stages, ie: hydrolysis, acidogenesis, acetogenesis, methanogenesis, etc., with main functions of specific microorganisms. Therefore, 16s rRNA sequencing analysis of microbial community was implemented at the eight-time points (Figure 4-3) representing different process stages, which aims at comparing microbial population changes in the anaerobic treatment process.

3.3 Microbial community adaptation in anaerobic digestion

To determine the temporal response of the microbial population to wastewater loading, the microbial community of the eight different process stages (Figure 4-4) were characterized by high-throughput sequencing. For characterizing the microbial community in response to increasing wastewater loading, principal coordinates analyses with a Bray–Curtis dissimilarity matrix was calculated as a measure of time gradual shift of population structure among the loading period. A general trend revealed by microbial community analysis was the gradual change in population structure for one-year loading of wastewater, which was plotted by two-dimensional graphs. The distinct distance between T1 and T2 indicates that the microbial communities in the two-time points changed considerably. The survived population may have strong adaptability of biofuel wastewater loadings, while decayed microbial could not be able to confront the environment. However, the differences among T2 to T8 time points become not significantly, which suggests the minor shift of the microbial communities during the gradual wastewater loading (Figure 4-4). It may further signify the higher adaptive capacity of the survived microbial community in T2.

Correspondingly, estimates of population diversity, including Simpson, Shannon and Inverse Simpson, showed increasing trends from Stage 1 to 2, while the trend does not differ during the other stages (Table 4-3). These changes observed suggest that decreased diversity might be associated with the microbial ability to overcome wastewater perturbation caused by toxicity. It could also be concluded that only a portion of microorganisms is more likely to adapt itself to the anaerobic and toxic environment than others.

3.4 Microbial populations with a significant response to increases of wastewater loading

The changes in microbial community structure in response to wastewater loading was assessed from the phylum level (Figure 4-5). A significant change between T1 and T2 was found and could be mainly linked to the increased microbial consortium in *Thermotogae*, *Chloroflexi*, and *Cloacimonetes* in T2, while the decreased population in *Synergistetes*, *Spirochaetes*, and *Firmicutes*, and *Bacteroidetes*. However, the three increased phyla together accounting for greater than 60% of the microbial community in the later loading episodes. Thus, the decreased community diversity with some decreased population community observed after T2 in this study likely resulted in decreased functional diversity and redundancy, which subsequently signify the capacity of some bacterial community are not able to cope with unfavorable process conditions such as the transient addition of COD due to wastewater loading. In contrast, the considerably increased microbial consortium in T2 may suggest its strong capability to survive and thrive in the anaerobic wastewater feeding condition. It could be accused of some specific functions of these bacteria.

Despite the considerable changes in the community between T1 and T2, 85 SVs were identified through the seven stages, suggesting that these populations represented a core community for the anaerobic digestion of toxic wastewater. Therefore, we take an in-depth look into the abundant bacteria found during the treatment. The relative abundance higher than 0.5% was extracted (Figure 4-6). The most abundant populations at the SV level affiliated with SV100 from order *SJA-15*, class *Anaerolineae* and phylum *Chloroflexi*. Its relative abundance increased from 0.05% in T1 to 25.7% in T2, and with an increasing trend to 39.5% in T3. It was also found in other anaerobic studies, but the function is still unknown. However, *Chloroflexi* may have a crucial role in the granulation in methanogenic reactors due to their filamentous morphology (Bovio et al. 2019, Lee et al. 2018). A similar trend was exhibited by SV6 *Geobacter*, with the relative abundance rising from 3.8 % in T1 to 26.6 % in T2, and sustained at 17.9% in T7. In T8, another *Geobacter* SV207 increased from less than 1% in T4 to 13.8% in T8. It has been identified that *Geobacter* and *Methanosaeta* can stoichiometrically convert ethanol to methane (Rotaru et al. 2014). In addition, SV436 from genus *Candidatus Cloacimonas*, family *Cloacimonadaceae*, is sustained along the loading period. It is known (or suspected) to be anaerobic mesophilic acetogens (Lee et al. 2018). However, some other SVs existed, such as SV14 from family *Aerolineaceae*, S50 wastewater-sludge group from family *Rikenellaceae*, SV633 from order *Bacteroidales*, SV13 from family *Synergistaceae*, existed. However, the function of these bacteria remains unknown since it can not be classified from a finer level. The similarities in population response to process conditions suggest common functional niches between these bacteria.

3.5 Methanogens involved in the anaerobic conversion of biofuel process wastewater

As methanogens as members of the archaea are among the microbial populations most sensitive to process disturbances in anaerobic digestion (Chen et al. 2012), the responses of archaeal microbial populations during the increased wastewater loading were further studied by 16S amplicon analysis to understand the responses of these populations to toxic wastewater loadings.

16S amplicon sequencing analysis identified 24 archaeal SVs involved in the anaerobic digestion process. 11 out of 24 SVs persisted throughout the anaerobic treatment period, representing more than 99% of the archaeal abundance in the wastewater loading archaeal community. The most abundant methanogen populations included *Methanosaeta*, *Methanobacterium*, and *Methanomassiliicoccus* (Figure 4-7). The diverse existence of methanogens suggested the diverse methanogenesis function in the anaerobic process.

The most abundant methanogen is *Methanobacterium* ($79.3\% \pm 5.6\%$), suggests the primary function of hydrogenotrophic methanogenesis procedure in the reactor. It is interesting to find that 11 SVs belong to *Methanobacterium* existed, while these SVs show distinct phylogenetic distances among each other (Figure 4-8). However, these *Methanobacterium* show distinct distance with *Methanosaeta* and *Methanosiliicoccus*. *Methanobacterium* could be the possible methanogen in transforming formaldehyde to methane since by a central metabolic intermediate of methylotrophic metabolism can help assimilation, dissimilation, and detoxification of large quantity of formaldehyde (Escalante-Semerena and Wolfe 1984). The persistence of *Methanobacterium* in AnMBR of this study raises the possibility that *Methanobacterium*-like organisms may function in a broader range.

Methanosaeta accounted for $12.7\% \pm 3\%$ of archaea, which mainly was accounted by SV27. It is obligately anaerobic, gram-negative and non-spore forming rod with flat ends, and require acetic acid as the sole source of energy with the production of equimolar amounts of CH_4 and CO_2 as the metabolic end products. It suggested the existence of acetoclastic methanogenesis function in the reactor. In accordance, acetic acid was a main organic compound in the wastewater loading, while it is the main precursor of methane formation by utilizing the acid as a substrate. However, *Methanosarcina* was little detected in the AnMBR, could be attributed to the recalcitrant and toxic compounds (Ng et al. 2016) present in the lignocellulosic wastewater.

Methanomassiliicoccus was on a steady increasing trend during the loading stages, from less than 1% in T1 to 16% in T8. *Methanomassiliicoccus* can produce methane by reducing methanol with hydrogen as the electron

donor, and it is not able to produce methane when hydrogen or methanol only present (Dridi et al. 2012). Indeed, methanol is one of the most important organic compounds in the wastewater loadings.

3.6 Syntrophic populations potentially involved in the anaerobic conversion of biofuel process wastewater

While methanogens were the most prominent members of methanogenesis in the anaerobic digesters analyzed in this study, they are also closely associated with syntrophic bacteria that can oxidize VFAs (Mosbaek et al. 2016). In particular, a diverse of syntrophic bacteria continuous identified along with the wastewater loading episode, including *Syntrophaceticus*, *Syntrophobacter*, *Syntrophomonas*, *Syntrophorhabdus*, and *Syntrophus* (Figure 4-9).

Syntrophaceticus has the acetate-oxidizing ability to produce methane in coculture with hydrogenotrophic methanogens (Westerholm et al. 2010). *Syntrophobacter* can degrade propionate only in coculture with an H₂-using organism (Boone and Bryant 1980, Sedano-Nunez et al. 2018), while propionic acid was found in the wastewater loading. *Syntrophomonas* is a syntrophic butyric-oxidizing bacterium, which produces hydrogen and/or formate from the high potential electron donor, and is the thermodynamically unfavorable production of. It is capable of metabolizing butyric acid to acetic acid by the beta-oxidation pathway (Crabble et al. 2016, McInerney et al. 1979), while butyric acid was a component in the wastewater feedings. *Syntrophorhabdus* was proposed to convert aromatic compound to acetate and hydrogen in associated with hydrogenotrophic methanogens (Leven et al. 2012). Various types of aromatic compounds persisted in the feeding wastewater, such as 2-Methoxyphenol, 2-Methoxy-4-methylphenol, phenol, 2-Methylphenol, 4-Methylphenol, 3,5-Dimethylphenol, 3-Propylphenol, 3-Ethylphenol. Thus, the presence of syntrophy-related bacteria could be responsible for the existence of organic compounds accompanying possible methanogens resulting from wastewater overloading.

4. Conclusion

AnMBR is an efficient and sustainable option to convert recalcitrant biofuel process wastewater to renewable energy, methane. However, process unstable resulting from toxic wastewater loading has been one of the obstacles to the application of anaerobic digestion technology. Therefore, it is important to develop a pre-treatment processes and operation strategies (feeding and acclimation) that can allow a group of microbial populations to thrive and facilitate the biological conversion of biofuel organics on anaerobic digestion processes. Using AnMBR treating recalcitrant wastewater as the model system, results from this study show that the microbial community thrived from gradually wastewater loading acquired population diversity, suggesting the development of enhanced functional diversity of some specific bacteria potentially beneficial to overcoming subsequent occurrences of higher substrate loading.

With the gradual increase in organic loading of the process wastewater, the AnMBR was able to remove >90% of COD loading with stable biogas production, which coincided with the dominance of *Methanobacterium*, suggesting the functional importance of hydrogenotrophic methanogenesis in the anaerobic metabolism of the wastewater organics. Other methanogen populations, including *Methanosaeta* and *Methanomassiliicoccus*, were also found at significant levels in the AnMBR, indicative of the need for functional diversity to support the degradation of diverse organic constituents in the process wastewater. More importantly, the prevalence of multiple syntrophic bacteria (*Syntrophaceticus*, *Syntrophobacter*, *Syntrophomonas*, *Syntrophorhabdus*, and *Syntrophus*) likely enhanced the degradation of some fatty acids during the process.

Appendix D: Figures and Tables

Table 4-1 Components in the 1L feed.

Components	Amount in the 1L feed
Soybean Oil (mL)	0.79
Beer (mL)	18.59
Starch (g)	0.99
Skim Milk (g)	3.52
Yeast Extract (g)	0.51
NaHCO ₃ (g)	1.68
KHCO ₃ (g)	2.00
Met Source AN (mL)	1.10
Granule Aid (mL)	0.24
Biofuels WW (mL)	20.00

Table 4-2 Organic compound concentration in AnMBR feed.

Organic compound	Compound concentration(ppm)	C content contributed to TOC(ppm)
Ethanol	470	245
Formaldehyde	386	154
Acetaldehyde	237	129
Methanol	295	111
Acetic acid	263	105
Furfural	65.9	41
5-Methyl-2-furancarboxaldehyde	44.7	29
Acetone	35.75	22
Propionic acid	38.5	19
2-Methyl-2-cyclopenten-1-one	10.9	8.17
2-Methoxyphenol	9.13	6.18
2-Butanone	8	5.3
Butyric acid	8.3	4.53
2-Methoxy-4-methylphenol	6.22	4.33
Phenol	5.42	4.15
Valeric acid	6.3	3.7
1-(2-Furanyl)-ethanone	5.31	3.48
2,3-Dimethyl-2-cyclopenten-1-one	3.93	3
2,3,4-Trimethyl-2-cyclopenten-1-one	1.94	1.5
2-Methylphenol	1.56	1.21
4-Methylphenol	1.48	1.15
3,5-Dimethylphenol	0.93	0.73
3-Propylphenol	0.35	0.28
3-Ethylphenol	0.19	0.15
Hydroxyacetone	—	—
3-Methyl-2-cyclopenten-1-one	—	—
Hydoxyacetaldehyde	—	—
iso-Propanol	—	—

Table 4-3 Microbial community diversity during an episode of wastewater loading.

Time Point	Shannon	Simpson	Inverse Simpson
T1	3.59	0.94	17.38
T2	2.83	0.85	6.69
T3	3.03	0.88	8.36
T4	2.96	0.86	6.97
T5	2.91	0.83	6.00
T6	2.88	0.84	6.14
T7	2.97	0.86	7.29
T8	2.58	0.81	5.27

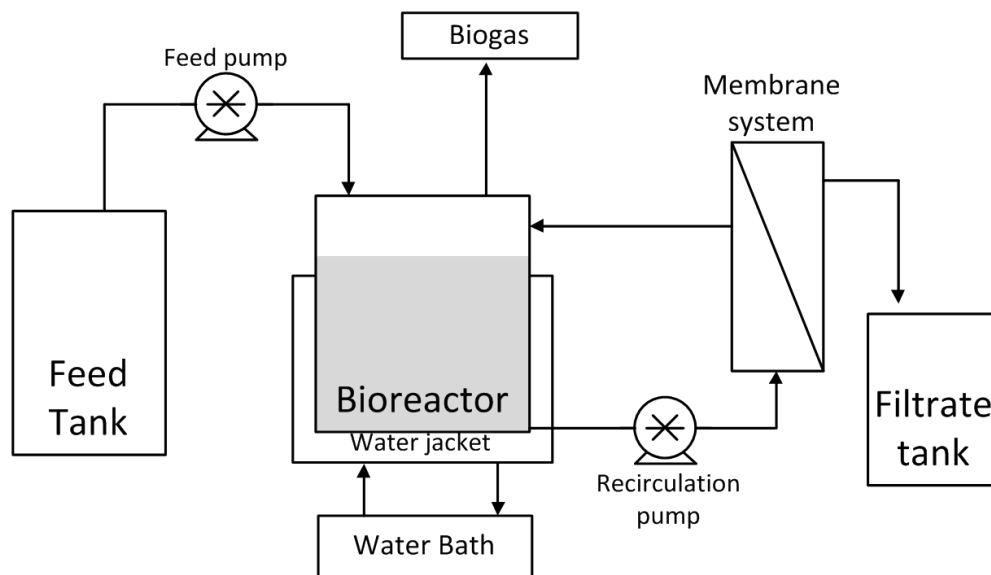
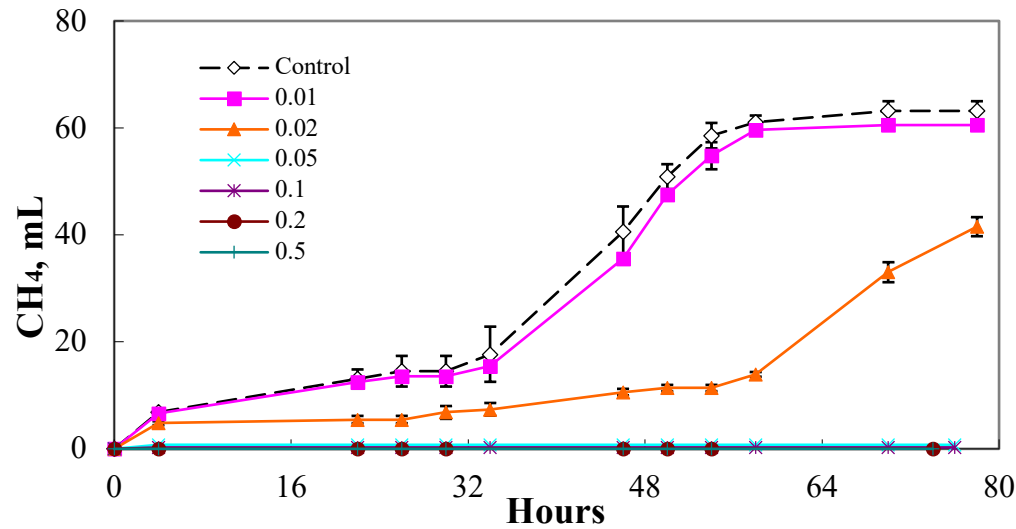


Figure 4-1 Anaerobic membrane bioreactor (AnMBR) system design.

(a)



(b)

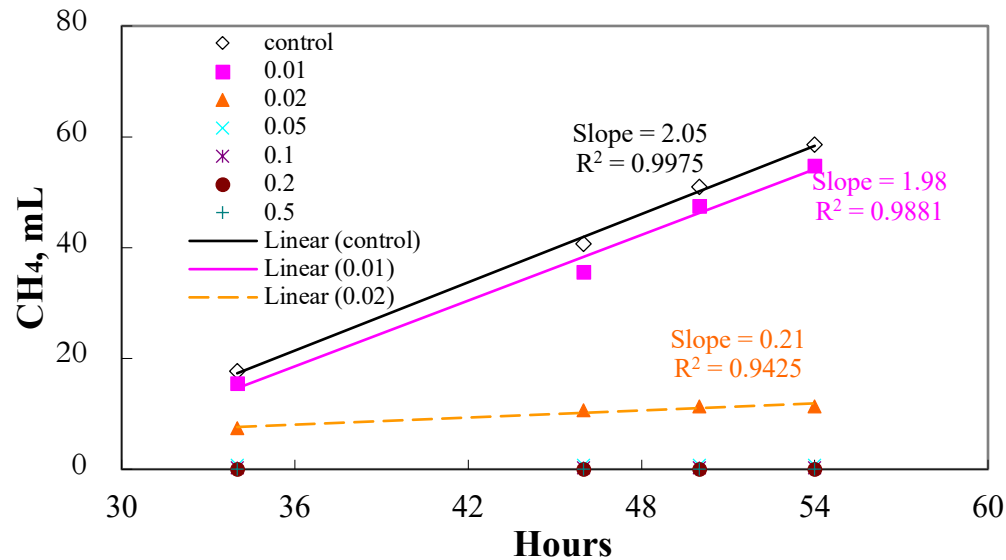


Figure 4-2 Pre-test of the biofuel wastewater.

(a) Accumulative methane, CH_4 production from pre-test digesters with biofuel wastewater at different loading rates. Colorful lines in the plot stand for the different loading rate of biofuel process wastewater. Control means no adding of biofuel process wastewater. 0.02 means 0.02% (vol/vol), the biofuel wastewater was diluted with the biodegradable organic solution as a volume ratio of 98 (nutritious feed): 0.02 (biofuel wastewater), others likewise. (b) Methane production rates of batch reactors with different loadings of biofuel process wastewater during the exponential growth phase.

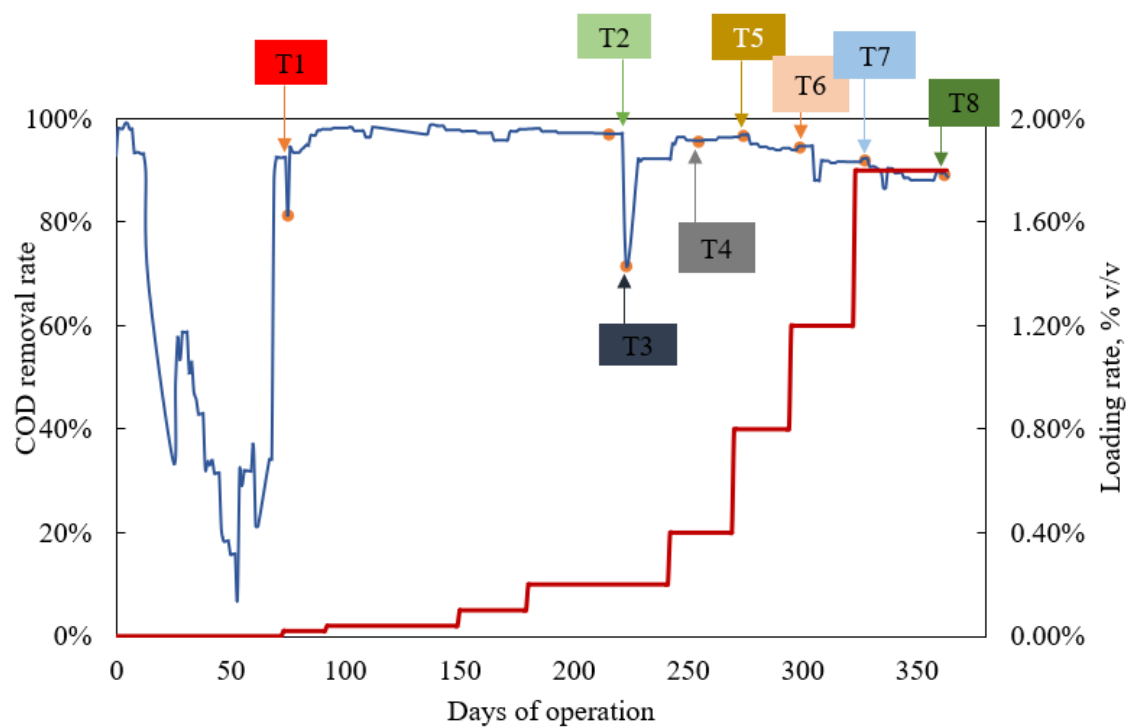


Figure 4-3 Variation of COD removal rate and wastewater loading rate (% v/v) in anaerobic digestion.

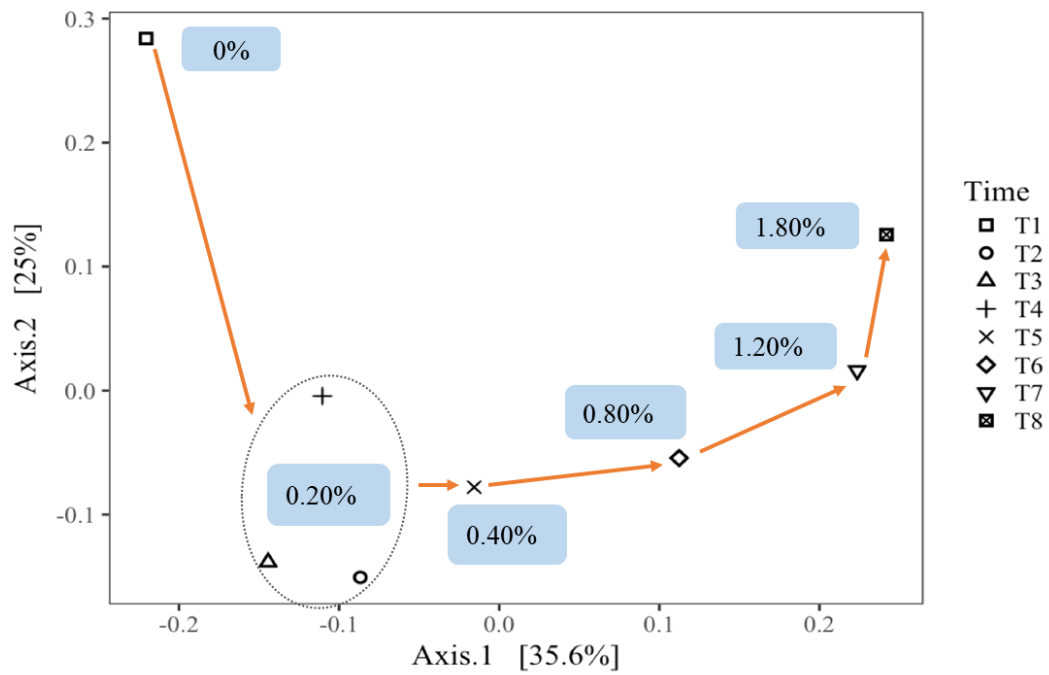


Figure 4-4 Principal coordinates analyses of the samples.

The analysis was performed using a Bray–Curtis dissimilarity matrix that described the time gradual shift of microbial community. The percentage label stands for % vol/vol.

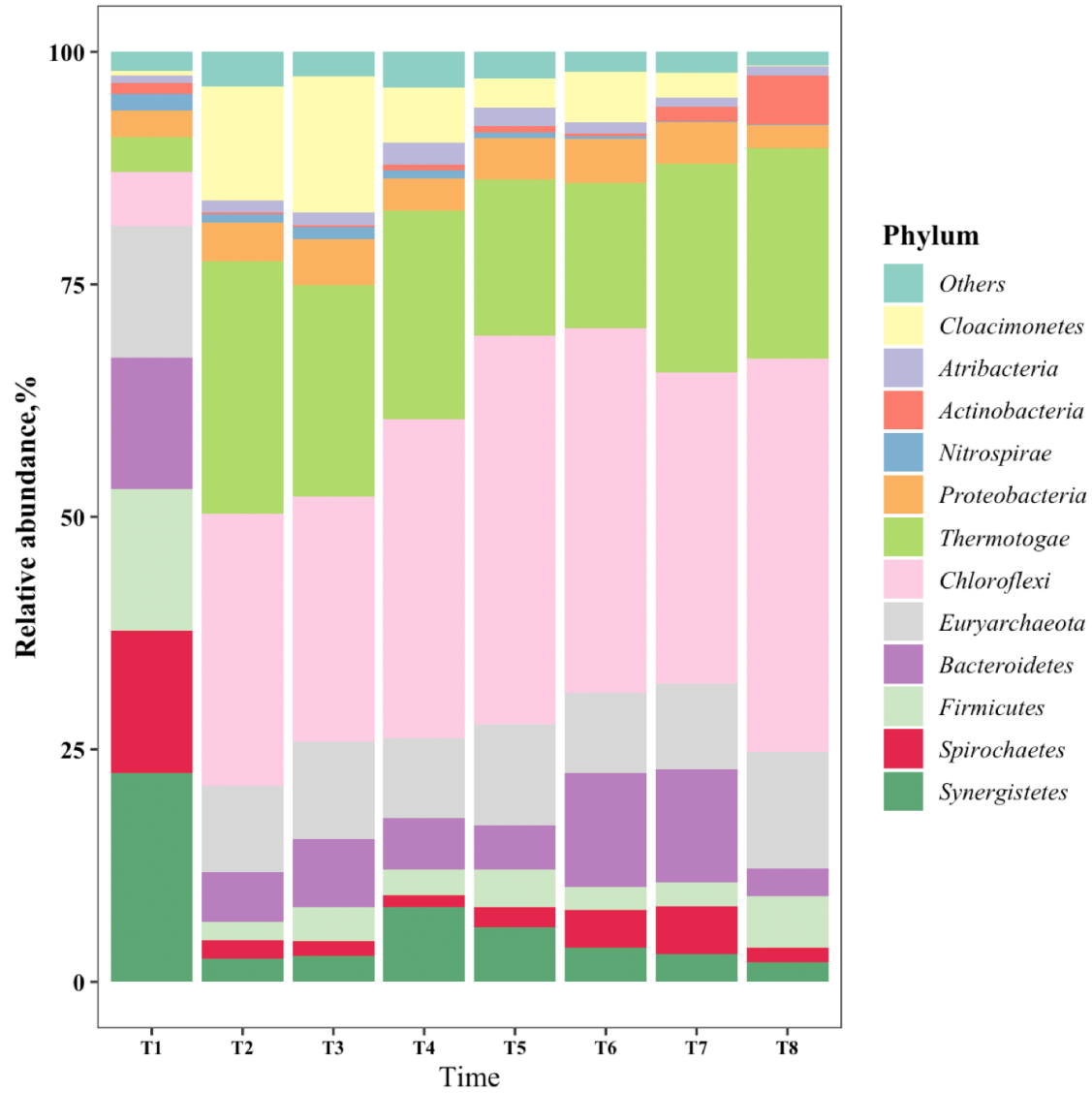


Figure 4-5 Microbial community at phylum level in anaerobic digestion.

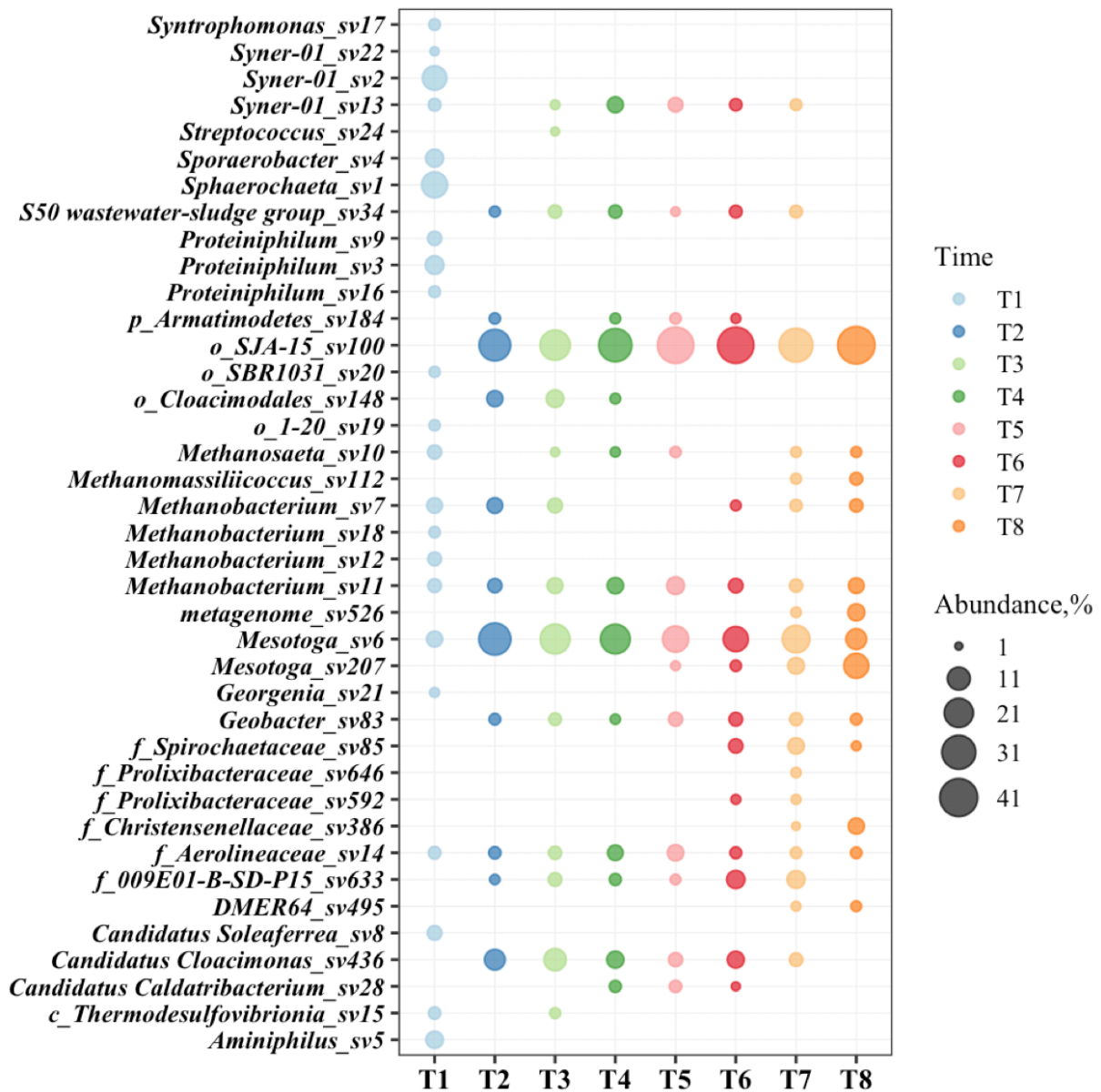


Figure 4-6 Abundant bacteria with relative abundance greater than 0.5% found in any sample during anaerobic treatment.

Taxonomy labels were from the genus level, if this bacteria cannot be classified from the genus level, it was labelled at a higher level. eg. *p_Armatimodetes_sv184* stands for sequence variant (sv) 184 from phylum Armatimodetes. *f_*, *o_*, *c_*, *p_* respectively stands for family, order, class, phylum level classification.

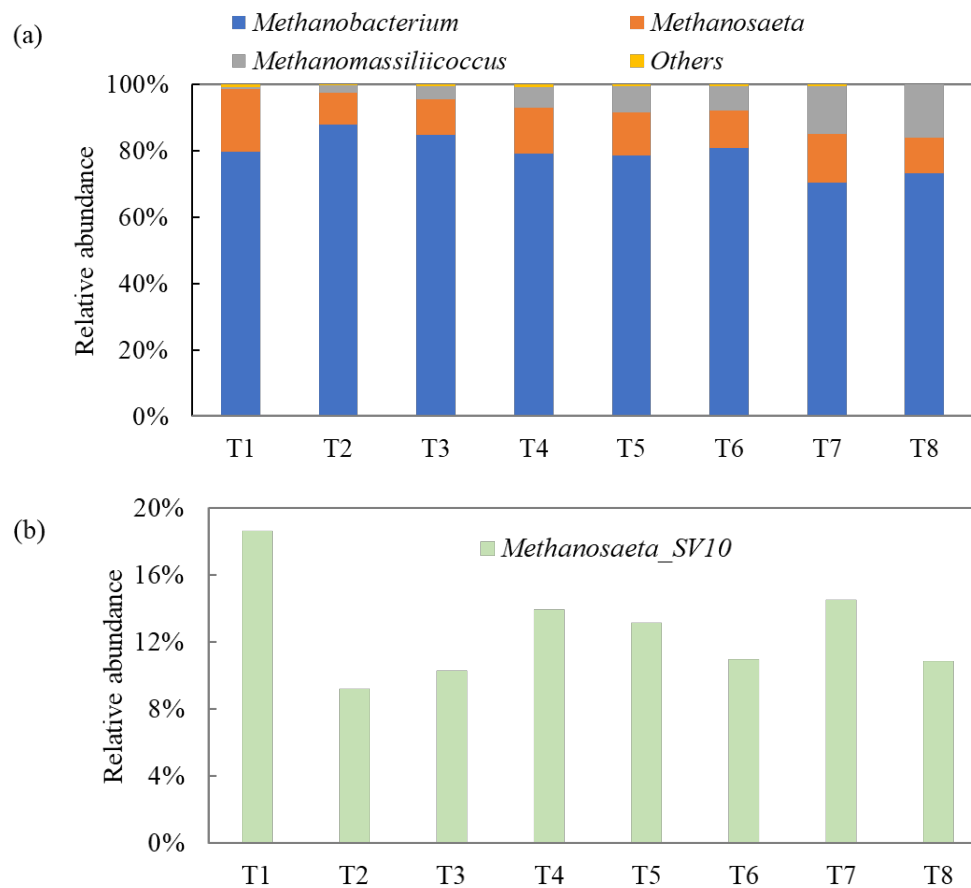


Figure 4-7 Microbial community change in anaerobic digestion.

(a) The relative abundance of identified methanogen populations as determined by 16S rRNA gene sequences in anaerobic digestion. (b) Gradually change *Methanosaeta* sequence variant detected in the digester. (c) Gradually change of *Methanomassiliicoccus* sequence variant found in the digester. (d) High abundant *Methanobacterium* sequence variants identified during wastewater loading.

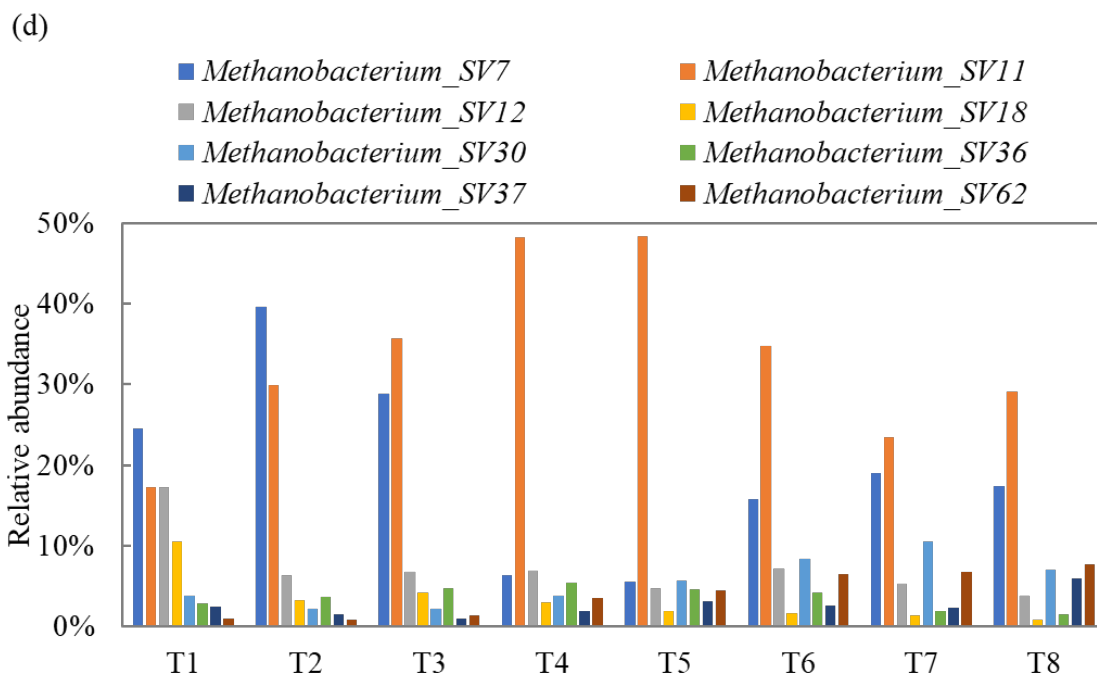
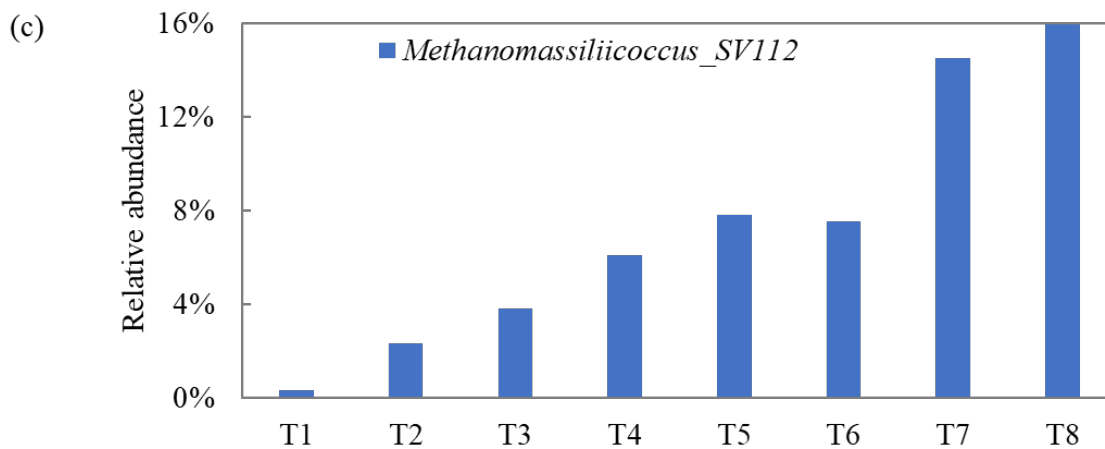


Figure 4-7 Continued

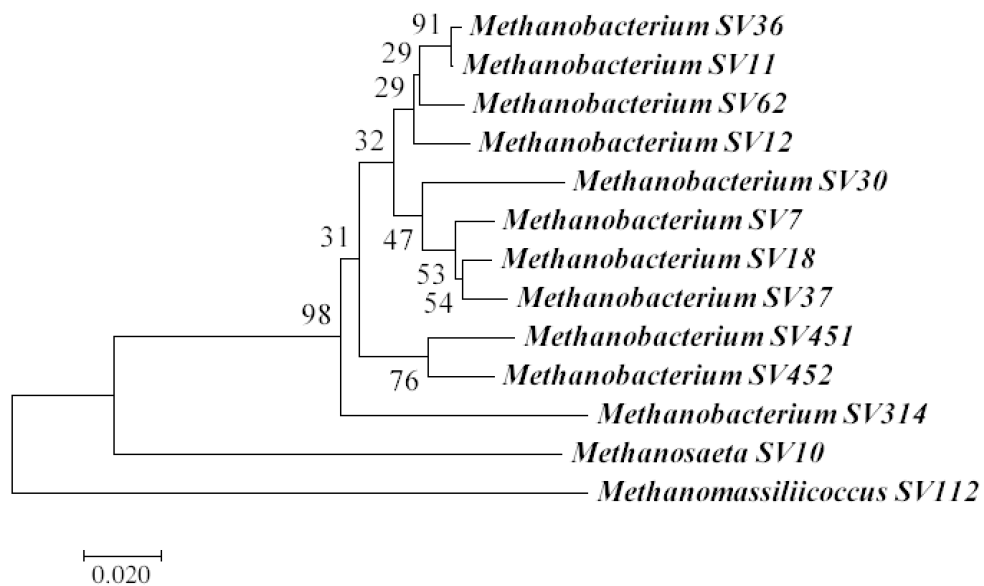


Figure 4-8 Neighbor-joining phylogenetic tree showing the relationships of 16S rRNA gene sequences of *Methanobacterium*, *Methanosaeta*, and *Methanomassiliicoccus*.

The numerical values at branch nodes indicate bootstrap values per 100 resamplings. The scale bar represents the number of substitutions per sequence position.

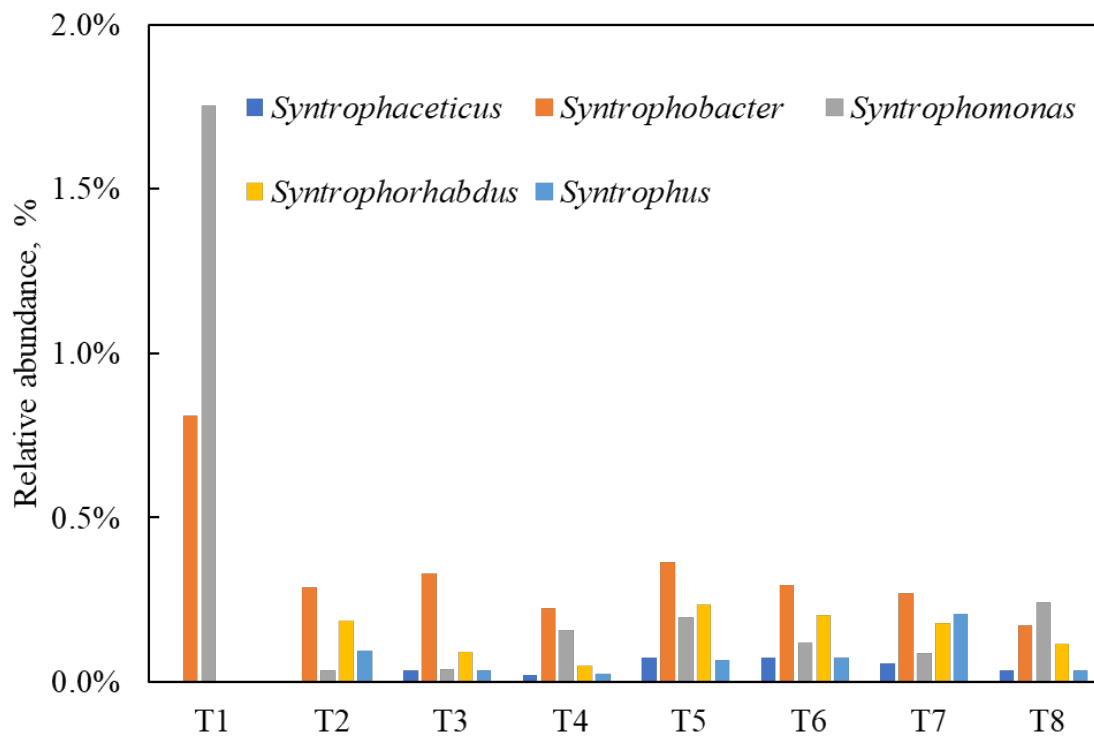


Figure 4-9 Possible utilization of influent organic compounds by syntrophic bacteria.

CHAPTER 4
POLLUTANTS SOURCE TRACKING IN NATURAL PROCESS

Section A: Rapid Change of Microbiome in Raw Sewage

Abstract

Sewage, an excellent representation of human fecal microbiome, is a pollution source of the urban waterbody. It is a widely used component of the source library in current source tracking study. It has been found that the human fecal microbial population can significantly decay when it enters into the natural environment, due to the low survival ability of these microbial community. Therefore, we hypothesized that fresher sewage could be a better microbiome assemblage in microbial source tracking. To investigate sewage microbiome assemblage shaped at different stages, 16S rRNA sequences from the pump station, WWTP influent, and 24-hour composite samples were detailed compared. Results showed that a subset of human fecal microbiome rapidly decayed in raw sewage, a group of non-fecal microbiome propagated in sewage conveyance system, and a group of human fecal microbiome thrived by sewage aging. Greater microbiome diversity arising with the aging and transportation of sewage in the natural environment, which reshapes the community composition of sewage. Non-overlaps of network hubs, nodes, and connectors were found among constructed RMT networks. Our results suggest that sewage from pump station and influent could be better microbiome pollution source in source tracking studies. And the persistence of some human fecal microbiome can be good indicators of newly polluted sewage in urban waters.

Keywords: sewage; microbiome community; Lefse; network analysis; source tracking

1.Introduction

More than 1 million miles of sewer lines, 16,000 wastewater treatment facilities are operating across the United States. However, sewer lines leakage is commonly found due to the blockages, sewer breaks, improper construction or other reasons beyond the control. Such an infrastructure problem increases the risk of public health and aquatic organisms by way of exposure to sewage-contaminated natural water. Sewage contamination can cause a chronic problem, which is expressed in high abundance of commensal organisms in the urbanized area (McLellan and Eren 2014). Therefore, it is of great importance to address the microbiome pollution source of sewage.

Microbiome source tracking combined with next-generation sequencing techniques has been recognized as an important technique in identifying pollution sources. Sewage is one of the most important pollution sources involved. To date, most studies have been focused on influent wastewater, treated wastewater, activated sludge. *Bacteroidales* (Haugland et al. 2010), *Lachnospiraceae* (Newton et al. 2011), and *Clostridiales* microbial communities have been constantly reported across different wastewater treatment plants influents (McLellan et al. 2010a). A group of core microbial community has been found in the activated sludge samples, which includes *Ferruginibacter*, *Prosthecobacter*, *Zoogloea*, *Subdivision3_genera_incertae_sedis*, *Gp4*, and *Gp6*, etc.

However, little has been emphasized on the microbial composition difference of different types of untreated sewage. Sewage has been identified as a reflection of the human gut microbiome. Most human gut microbiomes are anaerobic, which are not able to survive in the natural environment for a long time. Only 10%~15% of sewage microbiome was found as human fecal origin under the comparison of sewage samples and human fecal samples from the Human Microbiome Project (Newton et al. 2015). Fecal indicator bacteria and human-associated markers in the wastewater show a decaying trend in a different environment, such as beach sand, seawater (Zhang et al. 2015a). Therefore, microbiome composition in sewage, a capture of the human fecal microbiome, may suffer changes after it was discharged and transported to wastewater treatment plants (WWTPs). Soil, greywater, and stormwater would import into the sewer system (McLellan et al. 2010b), which contribute to the community diversity in the sewer system. But how it does it change the sewage microbial community remains unknown. Recently *Arcobacter*, *Acinetobacter*, *Aeromonas*, and *Trichococcus* have been identified as a propagated microbial community in the sewer system by the comparison between sewage samples and sewer biofilm samples (McLellan and Roguet 2019). But these samples have not considered the impact of aging in a short period of time since the long transportation time among geographical sites is inevitable.

In this research, we combined insights from 16S rRNA genes in three types of sewage and human fecal microbiome to identify taxonomy at the community level. We use sewage samples at different stages from the same treatment plants and human fecal microbial data sets to test the hypothesis that different type of sewage has its own assemblage. We evaluated whether microbial community change with its process variables and how the microbial community changes with the transportation and time effect. RMT network analysis and Lefse analysis were performed to figure out the keystone taxa. Together, these analyses aimed to identify the most suitable sewage type for microbial source tracking.

2.Methods

2.1 Sampling sites selection and sample collection

Sewage samples were collected from 6 wastewater treatment plant (WWTP) sites from across Knoxville, TN. Geographically, wastewater treatment plants and wastewater pumping stations distributed in and around Knoxville have been selected (Figure 5-1), which consists of Kuwahee wastewater treatment plant, Loves Creek wastewater treatment plant, Fourth Creek wastewater treatment plant, First utility wastewater treatment plant, Oak ridge wastewater treatment plant, London utility wastewater treatment plant, Lenior utility wastewater treatment plant. Pumping stations include three pump stations respectively connected with the Kuwahee wastewater treatment plant, Love Creek wastewater treatment plant, and Fourth Creek wastewater treatment plant.

Sewage samples include 24 hours of sewage samples from multiple wastewater treatment plants, wastewater treatment plant pumping stations, and septage tanks. Composite samples and influent samples were collected according to each wastewater treatment plant's standard collection procedures, which including single-time-point grab samples to flow-weighted composite samples of 24 hours running time. Pump station samples were collected in each WWTP's pump station, and samples were collected by one-time-grab from pump outlets. For better capture residential wastewater fingerprint, all one-time-grab samples were selected at 8:00 am. For excluding the stormwater influence on sewage, at least 72 dry hours were ensured before wastewater sampling day. Following sample collection, we transferred the samples to autoclaved Nalgene bottle and then placed the sample bottle in cooler with ice, and samples were transported to the lab within 4 hours.

Sewage samples were mixed by shaking, and the microbial communities were collected by filtrating of enough wastewater onto membrane filters (Whatman glass microfiber filters 934-ah). For making sure of enough quantity and diversity of microbiome tracked, enough sewage was filtered until sewage cannot pass through the filter. All filters were stored in a 2-ml centrifuge tube at -80°C until DNA extraction. In addition, environmental parameters, including 5-day biochemical oxygen demand (BOD₅), pH, and total suspended solids (TSS), BOD₅ were measured. were measured by YSI Oxygen Meter and YSI stirring bottle probes according to the standard methods (APHA, 2005). TSS and pH were also measured according to the standard methods.

2.2 DNA extraction and 16S rRNA sequencing

DNA extractions were performed using MP FastDNA™ Spin Kit for Soil according to the manufacturer's instruction (Protocol Revision #116560200-201608). Briefly, the biomass samples were resuspended in the 978 µl Sodium Phosphate Buffer solution and 122 µl MT Buffer, then followed by 100 seconds vortex on MP biomedical fast prep instrument. Then the DNA was purified with Zymo Genomic DNA Clean & Concentrator™-10 kit according to the instructions provided by the manufacturer. The DNA concentrations were determined using Thermo Scientific NanoDrop ND-3300 Fluorospectrometer.

Polymerase chain reactions were prepared with a cocktail mix containing 12.5 µL Phusion flash Master Mix, 10 µL ultra-pure water, 1 µL forward primer, 1 µL reverse primer, and 2 µL (100 to 150 ng) DNA template. Each universal primer pairs were designed to target the V4 region of bacterial and archaeal 16S rRNA genes: forward primer 515F (GTGCCAGCMGCCGCGGTAA) with 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), and forward primer link (GT); reverse primer 806R (GGACTACHVGGGTWTCTAAT), with adapter (CAAGCAGAAGACGGCATACGAGAT), link (CC), and a unique 12-base specific barcode for each sequence. The PCR program included one cycle of 94 °C for 3 minutes, followed by 35 cycles at 94 °C for 45 seconds (denaturation), 55 °C for 1 minute (annealing), and 72 °C for 1 minute and a half (elongation), a final extension at 72 °C for 10 minutes, and then stored at 4 °C. After the program finished, the samples were run on an Agilent 2100 Bioanalyzer Instruments with Agilent DNA 7500 chips to ensure the amplicon quality and measure the DNA concentrations. Post PCR, amplicons were pooled (a set of 9 or 10) based on the peak height from Agilent DNA 7500 analyses and the concentrations were measured using KAPA Illumina kit for quantifying library

(KK4824) from KAPA Biosystems. Samples were pooled again according to the results and number of sequences desired into a final one and the concentration was measured again with the KAPA quantifying kit. 4 nM pooled libraries were diluted with 10 mM Tris/0.05% tween buffer (pH = 8.5) and denatured with 0.2 N freshly made NaOH solution. The sample was then combined with 20% PhiX control kit to increase the library diversity. The mixture was incubated at 96 °C for 2 minutes using a heat block and then put into a water-ice bath for 5 minutes. After the sample was loaded, an Illumina MiSeq System with a MiSeq Reagent Kit v2 (300-cycles) and Metagenomic workflow was selected to execute the 16S protocol using the MiSeq Reporter software (MSR) at the University of Tennessee Genomics Core (UTGC).

2.3 Dataset construction and statistical analysis

We downloaded human fecal samples (Caporaso et al. 2011) from Earth Microbiome Project, which was also sequenced using the hypervariable region 4 (V4) 16S rRNA sequencing protocol. Therefore, 202 human fecal samples were included for comparison of the human fecal fingerprint and sewage samples.

Raw sequences data of human fecal samples and sewage samples were processed in QIIME2 DADA2 respectively, resulting in unique 786 sequence variants in human fecal samples and 8243 unique sequence variants in sewage samples. Taxonomic composition of these sequence variants was explored by a pre-trained Naïve Bayes classifier on Greengenes 13.8 database, where the sequences were trimmed to bound by the 515F/806R primer pair. Then we compared the sequence variants against the pre-trained classifier. Since human fecal sequences are in the single-ended format and sewage samples are in the pair-ended format, they were collapsed at genus/family level for comparison. Sequences were rarified by `rarefy_even_depth` in `phyloseq` R at 10000 sequences for later analysis. A principal coordinate analysis (PCoA) was performed to evaluate the dynamics of microbial population among different sewage microbiome and human fecal microbiome.

Phylogenetic trees were constructed from all representative sequences in QIIME2. In general, the pipeline uses the `mafft` program to perform a multiple sequence alignment of the sequences, then `FastTree` was applied to generate the phylogenetic tree from the masked alignment. Faith's Phylogenetic Diversity (Faith's D) was calculated thereafter under the sampling depth of 10000. The phylogenetic distance between sequence variants was determined by their relatedness in the phylogenetic tree in `picante` package in R. Also, other alpha diversity index, observed OTU, Chao1, ACE and Fisher index, were calculated in `phyloseq` package in R.

Alluvial plot was plotted in `ggplot` package in R, heatmap was plotted in `heatmap.2` package in R. Network analysis was performed in the network analysis pipeline at <http://ieg4.rccc.ou.edu/MENA/>. We constructed networks according to sewage type, pump station sewage, influent sewage, and 24-hour composite sewage. The networks constructed were after that named as pump, influent, and composite. In brief, sequence variants with prevalent among 80% of total samples were kept in the network construction, missing data were filled with 0.01 in blanks with paired valid values. Sequence variant counts were taken logarithm and analyzed by Pearson correlation coefficient (r). Random matrix theory (RMT) based algorithms were applied to determine the cutoff r value for all networks. Then global network properties, individual nodes' centrality, module separation and modularity calculation, power-law models fitting, randomize the network structures were also analyzed in MENA pipeline. Therefore, a group of feature set, total number of nodes, total number of edges, average degree (avgK), centralization of degree (CD), average cluster coefficient (avgCC), average geodesic distance (GD), centralization of betweenness (CB), centralization of stress centrality (CS), density, number of modules, modularity and R^2 of power law (Deng et al. 2012), were calculated. Detailed network characters were described in Table 4-2. The networks were visualized in Cytoscape software.

Multiple regression of distance matrices (MRM) was applied to evaluate the importance of community and environmental parameters for network topological feature in `ecodist` package in R. Network topological feature and environmental parameters were standardized with `decostand` function in `vegan` package in R. Euclidean distance matrices of two datasets were applied in the MRM models.

3.Results

3.1 Microbial community composition in sewage and human fecal samples

We collected 15 pump station sewage samples, 15 influent wastewater samples, and 40 24-hour composite sewage samples in the same season. In order to ensure the quality of the read, all samples were rarefied at 10000 depth, which retained 59 samples finally. We also collected 202 human fecal samples from EMP dataset. NMDS analysis of the microbial community provides estimates of sample relatedness among wastewater samples and human fecal samples, which showed that sewage samples behaved distinct different from human fecal samples. NMDS analysis further revealed that the bacterial community structure differed considerably among sewage samples (Figure 5-2). While three divergent clusters were found among three types of sewage samples, a minor decay distance of sewage samples to human fecal samples existed. Pump station samples showed the closest distance to human fecal, followed by the influent samples, while the 24-hour composite samples revealed the least one. The decay distance to sewage samples indicating that sewage may contain uneven richness and diversity in the microbial community.

Comparison of human fecal samples and sewage samples from the family level (relative abundance greater than 1% were retained) revealed a major microbial difference between human fecal samples and sewage samples (Figure 5-3). Among all abundant families, pump station samples got the highest abundance retained, influent ranked next to it, and composite samples were at the least level. It may suggest a more diverse bacteria existed in composite samples than pump station samples and influent samples. In comparison, seven families made up of nearly 98% of the human fecal microbiome, which suggests a low diversity of the microbiome. The most dominant taxa, *Bacteroidaceae*, accounted for more than 60% of the population. *Ruminococcaceae* and *Lachnospiraceae* respectively accounted for 20% and 15% of the total abundance.

Sewage exhibited the effect of a microbial mixture of a group of the human fecal microbiome, while sewage samples showed a decreasing trend in human fecal microbiome fingerprint. Microbiome commonly existed in human samples are typically most abundant in pump station samples, with more than 91% of human fecal contained families exist. Influent sewage samples got less than 89% of human fecal families, while composite samples were at the lowest level with 87% (Figure 5-4). It may imply that some human fecal microbiome decay in sewage transport. However, further alpha diversity analysis found that sewage showed an increasing trend in alpha diversity during transport (Figure 5-5). They have observed OTU, Chao1 index, ACE and Fisher index consistently explicit the highest diversity level in the composite sample, the lowest level in pump station samples. It could be attributed to the decay of specific bacteria but an increase of some other propagated bacteria. The human fecal microbiome was in lower diversity than sewage samples (McLellan et al. 2010a). Therefore, it is reasonable to find that pump station samples are at the lowest diversity compared to the other two types of sewage. The longer time sewage discharged in the natural environment; the more additional microbiome inputs will be.

Additionally, Pearson correlation analysis between environmental parameters (including BOD₅, TSS, VSS, pH) and diversity index was performed (Figure 5-6), which disclosed a positive relation between BOD₅ and alpha diversity index ($p < 0.001$), and BOD₅ has a significant positive relation with VSS ($p < 0.001$, $r = 0.64$). Therefore, the higher diversity of sewage microbiome, the higher BOD₅, and VSS of sewage.

3.2 Effects of transportation and aging on sewage microbiome

The microbial community in pump station and influent contains a unique group of bacteria that are distinct from the composite microbiome. As sewage was storage for 24 hours, fecal components decreased and became dominant by non-fecal microbiome (Figure 5-7). *Bacteroidaceae*, *Campylobacteraceae*, *Flavobacteriaceae*, *Prevotellaceae*, *Rhodocyclaceae*, *Weeksellaceae* and *Ruminococcaceae* were significantly found in pump station and WWTP influent, while little was found in composite samples. Among these families, *Bacteroidaceae*, *Ruminococcaceae*, and *Prevotellaceae* are the important human fecal-associated taxonomy. *Flavobacterium* was the only bacteria found in *Flavobacteriaceae* (belongs to *Bacteroidetes*), while a variety of species were found,

consisting *Flavobacterium succinicans*, *Flavobacterium frigidarium*, *Flavobacterium gelidilacus*, *Flavobacterium columnare*, and others. *Rhodocyclaceae* phylogenetically affiliated with the *Betaproteobacteria*, *Uliginosibacterium*, *Dechloromonas*, *Zoogloea*, *Azospira*, *Azoarcus*, *Dok59*, *Candidatus Accumulibacter*, *Sulfuritalea*, *Hydrogenophilus*, *Propionivibrio*, *Sterolibacterium*, K82, C39, *Thauera*, *Azovibrio*, *Petrobacter*, *Rhodocyclus* were found in sewage. *Weeksellaceae* belongs to *Bacteroidetes*, genus *Wautersiella*, *Cloacibacterium*, *Chryseobacterium*, *Elizabethkingia*, *Ornithobacterium*, *Riemerella*, *Weeksella*, and *Bergeyella* were frequently found in wastewater (Stadler et al. 2018). On the contrary, *Lachnospiraceae*, *Leptotrichiaceae*, and *Streptococcaceae* showed an increasing trend with the aging of sewage, which may suggest the enrichment or additive source of this microbiome. *Lachnospiraceae*, affiliated with Firmicutes, is one of the most abundant groups of fecal bacteria in sewage, while the distribution of human fecal *Lachnospiraceae* can be reflected by sewage. *Leptotrichiaceae* belongs to *Fusobacteria*, *Leptotrichia* was the most abundant genus found, it is a part of the microbiome in the oral cavity, human female genitalia, and intestine (Gao et al. 2016). Furthermore, *Moraxellaceae* and *Comamonadaceae* existed relative stable among three types of sewage *Acinetobacter*, and *Enhydrobacter* is the top-ranked two genera in *Moraxellaceae*, while *Acidovorax* and *Comamonas* abundantly exist in *Comamonadaceae*.

The microbial community in different types of sewage contains a unique assemblage of bacteria that are distinct from each other. Heatmap with hierarchical clustering by average linkage agglomeration method disclosed the genus distribution pattern among them (Figure 5-8). The significant different cluster was found between composite and influent/pump station samples, while pump station samples and influent samples also showed different patterns. Significant abundant of *Ruminococcus*, *Bacteroides*, *Prevotella*, *Zoogloea*, *Faecalibacterium*, *Coprococcus* and *Roseburia* were found in the pump station. Except for *Zoogloea*, others all belong to the human fecal origin. The extensive decay of some fecal bacteria suggests the unadaptability of some fecal originated bacteria. *Ruminococcus* is host-associated bacteria, while very few *Ruminococcus* are consistently found in the natural environment (La Reau et al. 2016). *Bacteroides* and *Prevotella* cannot survive for a long time since environmental conditions make it not adaptable.

Similarly, *Bacteroides* species and *Prevotella* species have been regarded as markers of recent fecal pollution since the survival period of *Bacteroides* was shorter than fecal coliforms and enterococci (Balleste and Blanch 2010), and they have little potential for growth or persist in the natural environment (Kreader 1998). *Faecalibacterium*, *Coprococcus*, and *Roseburia* are butyrate-producing bacteria, phylogenetically affiliated with *Clostridia*. *Faecalibacterium* only contains one species, *Faecalibacterium prausnitzii*; an anaerobe widely exists in the human gastrointestinal tract. However, little has been studied on the survival ability of *Coprococcus* and *Roseburia* in the natural environment.

However, when sewage was transported from the pump station to WWTPs, pump station abundantly existed bacteria decayed in a large degree, *Cloacibacterium*, *Arcobacter*, *Macellibacteroides*, *Flavobacterium*, *Streptococcus*, *Comamonas*, *Acidovorax*, and *Enhydrobacter* increased. Except for *Streptococcus*, the other three are all not human originated. *Cloacibacterium* was typically wastewater bacteria, specifically, the key functional bacteria of the anaerobic environment of textile dyeing industry wastewater (Zhang et al. 2017a). *Arcobacter*, planktonic bacteria, which can use nitrate as the terminal acceptor and tolerate to sediment condition. In our study, most *Arcobacters* belonged sequence variants are classified as *Arcobacter cryaerophilus*. It could be caused by the water temperature; all sewage samples were sampled in winter. Cold water seems to favor *Arcobacter cryaerophilus* compared with other species of *Arcobacter* (Levican et al. 2014). *Macellibacteroides*, the only species *Macellibacteroides* fermentans found in sewage, is a novel obligately anaerobic, non-spore-forming, rod-shaped mesophilic bacterium, but little has been studied on it. *Comamonas* is versatile in degrading aromatics. It is also a versatile denitrifier (Ma et al. 2015). The rising relative abundance in influent sewage could be related to increasing concentration of aromatics or increasing of concentration of organic compounds. *Acidovorax*, an important denitrifying microbial community in WWTP (Lim et al. 2005), was formerly found as dominant bacteria in the anaerobic manhole. *Enhydrobacter* is rarely studied, which only found in primary influent sewage (Shanks et al. 2013).

Furthermore, 24-hour composite samples contained significantly increased genus in *Acinetobacter*, *Leptotrichia*, *Lactobacillus*, *Blautia*, *Bifidobacterium*, *Lactococcus*, *Collinsella*, *[Ruminococcus]*, *Paracoccus*, and *Rhodobacter*. Most of these bacteria are human originated bacteria, which is consistent with former literature. It identified that human gut microbiome could survive, grow, and reshape a new assemblage of sewage environment. *Blautia* has been found as an ideal human fecal pollution biomarker, which can evaluate the amount of human influence (Koskey et al. 2014). *Bifidobacterium*, especially *Bifidobacterium adolescents*, has been used as human-associated markers in microbial source tracking (Jeanneau et al. 2012), which more related to recent sewage contamination (Carrillo et al. 1985). *Lactococcus* is in low abundance in human fecal samples, while is enriched in 24-hour samples, which indicates the natural aerobic environment may provide advantages to their adaptation and survival (Liu et al. 2015).

Also, Lefse analysis was performed by first analyzed all sequence variants under Kruskal-Wallis test, while sequence variants violated the null hypothesis are further analyzed by the pairwise Wilcoxon test to check all pairwise comparisons between subclasses within different classes significantly agree with the class level trend. Finally, the resulting subset of vectors was used to build a Linear Discriminant Analysis model from which the relative difference among classes is used to rank the features (Segata et al. 2011). As a result, 23, 13, and 11 biomarkers were found in a pump station, wastewater treatment plant influent, and 24-hours composite samples (Figure 5-9). Interestingly, consistent with previous results, top abundant sequence variants respectively belong to *Bacteroides*, *Prevotella*, and *Acinetobacter* were found on a steady decreasing trend along with the transport and the aging of sewage (Figure 5-10). An increasing trend of sequence variants from *Flavobacterium*, *Cloacibacterium*, *Streptococcus*, and *Arcobacter* were found in influent but composite samples, which suggests the propagation of this microbiome in urban sewer conveyance system. Interestingly, the abundance of different *Bifidobacterium* was uniquely higher than pump station sewage and influent sewage, which suggests the consistent survival characteristics of *Bifidobacterium*.

3.3 Analysis of network characteristics and community/process variables

Different sewage samples were categorized into a network, which resulted in pumping station network, influent network, composite network. Network topological features were shown in Table 5-1. The networks expressed scale-free, small world, and modularity, which is consistent with former results. The three networks fit the power law distribution very well, suggests some sequence variants can be connected well with others, indicating the scale-free character. Small-world behavior was observed in the constructed networks, which indicated by the significant different value of average geodesic distance (GD) and average clustering coefficient (avgCC) in empirical networks and random networks. For modularity, randomized networks showed lower modularity values than empirical networks, which suggested the modularity of networks.

Within-module connectivity (z_i) and among-module connectivity (P_i) define the topological roles of nodes. The within-module connectivity characterizes within module connectivity, while the among-module connectivity illustrates between module connectivity. Four classes, peripheral nodes ($z_i \leq 2.5$, $P_i \leq 0.62$), connectors ($z_i \leq 2.5$, $P_i > 0.62$), module hubs ($z_i > 2.5$, $P_i \leq 0.62$) and network hubs ($z_i > 2.5$, $P_i > 0.62$). Peripheral nodes represent specialists whereas the other three are generalists (Deng et al. 2012). Among the three networks, each network contains network hubs, module hubs, and connectors (Table 5-2). A total of 9 sequence variants were identified as module hubs (Figure 5-11), but none of them were classified as module hubs among any two constructed networks, which suggests the unique assemblages of each type of sewage. Likewise, none of the connectors identified overlaps among different networks, suggesting each sewage microbiome change dynamically. Most network hubs, module hubs, and connectors were from phyla *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*.

Furthermore, the correlations of microbial community composition/environmental variables with network topological features were performed by MRM analysis (Figure 5-12). In the MRM model, microbial population characteristics, including alpha-diversity index and phylogenetic diversity, and sewage environmental variables including BOD₅, pH, and TSS, were included in the analysis. Results showed that phylogenetic diversity, BOD₅,

and pH were the most critical variables correlated with phylogenetic diversity. Moreover, network topology (density and transitivity) correlated with BOD₅, pH, and alpha diversity index (Fisher index).

4. Conclusion

Sewage from the pump station, WWTP influent and 24-hour composite samples were collected, while human fecal microbiome samples were downloaded from Earth Microbiome Project. 16S rRNA technique was applied in identifying the microbiome assemblage differences in sewage. Results showed that different types of sewage receive a different proportion of microbiome mapping from human fecal samples, which suggests the possible decay of human fecal population in sewage. The principal component analysis also revealed the microbiome structure difference among human fecal samples and different sewage samples. Microbial community change was also expressed in alpha diversity difference, which showed the increasing diversity from the pump station to 24-hour composite samples. The diversity index showed a significant positive correlation with BOD₅, while BOD₅ significantly positive correlated with VSS. The diversity change was mostly expressed in the decay of some human fecal microbiome from the pump station to influent, including *Ruminococcus*, *Bacteroides*, *Prevotella*, *Zoogloea*, *Faecalibacterium*, *Coprococcus*, and *Roseburia*. However, accompanied by decreased microbiome, *Cloacibacterium*, *Arcobacter*, *Macellibacteroides*, *Flavobacterium*, *Streptococcus*, *Comamonas*, *Acidovorax*, and *Enhydrobacter* increased. Furthermore, 24-hour composite samples got increased abundance of the genus in *Acinetobacter*, *Leptotrichia*, *Lactobacillus*, *Blautia*, *Bifidobacterium*, *Lactococcus*, *Collinsella*, [*Ruminococcus*], *Paracoccus*, and *Rhodobacter*. Such pattern was further confirmed by Lefse analysis, which showed that top abundant sequence variants from *Bacteroides*, *Prevotella*, and *Acinetobacter* were identified as biomarkers in pump station samples, *Flavobacterium*, *Cloacibacterium*, *Streptococcus* and *Arcobacter* in influent, and *Bifidobacterium* in 24-hour composite samples. Finally, RMT networks were constructed for each type of sewage. Non-overlap was found among networks, which suggests the independent assemblage of the microbiome in sewage at different stages. Furthermore, in coincide with former results, the network topology was found associated with BOD₅ and phylogenetic diversity.

Appendix E: Figures and Tables

Table 5-1 Topological features of the empirical molecular ecological networks in comparison to random network.

networks	Empirical networks					Random networks		
	Network nodes	R ² of power law	Average geodesic distance (GD)	Average clustering coefficient (avgCC)	Modularity	Average geodesic distance (GD)	Average clustering coefficient (avgCC)	Modularity
pump	179	0.922	4.99	0.104	0.701	4.149 ± 0.077	0.023 ± 0.009	0.572 ± 0.011
influent	157	0.855	4.267	0.258	0.486	3.149 ± 0.055	0.112 ± 0.017	0.401 ± 0.007
composite	112	0.662	3.27	0.254	0.481	2.924 ± 0.058	0.156 ± 0.020	0.392 ± 0.010

Table 5-2 Network hub, module hubs and connectors of microbial association networks.

Network	Hub	Network	Domain	Phylum	Class	Order	Family	Genus	Species
pump	sv42973	network hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	Unclassified
	sv44854	module hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Unclassified
	sv4357	module hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Paludibacter	Unclassified
	sv42873	module hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Unclassified	Unclassified	Unclassified
	sv50537	module hub	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Unclassified	Unclassified	Unclassified
	sv4282	connector	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Arcobacter	Unclassified
	sv32581	connector	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Proteocatella	sphenisci
	sv46237	connector	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	Vitreoscilla	Unclassified
influent	sv41716	network hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri
	sv40151	network hub	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Unclassified	Unclassified
	sv25892	module hub	Bacteria	Firmicutes	Clostridia	Clostridiales	[Acidaminobacteraceae]	Fusibacter	Unclassified
	sv6124	module hub	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas	diminuta
	sv38993	module hub	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	Unclassified
	sv32563	module hub	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	caccae
	sv38212	connector	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	minor
	sv46679	connector	Bacteria	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Propionivibrio	Unclassified
composite	sv52566	connector	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Unclassified	Unclassified	Unclassified
	sv41095	network hub	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	Unclassified
	sv15862	module hub	Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Leptotrichiaceae	Leptotrichia	Unclassified
	sv42197	module hub	Bacteria	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
	sv20134	connector	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Unclassified	Unclassified
	sv47034	connector	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	Unclassified	Unclassified
	sv28931	connector	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Agrobacterium	Unclassified
	sv2128	connector	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	Unclassified
	sv22664	connector	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megasphaera	Unclassified
	sv50462	connector	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas mexicana	

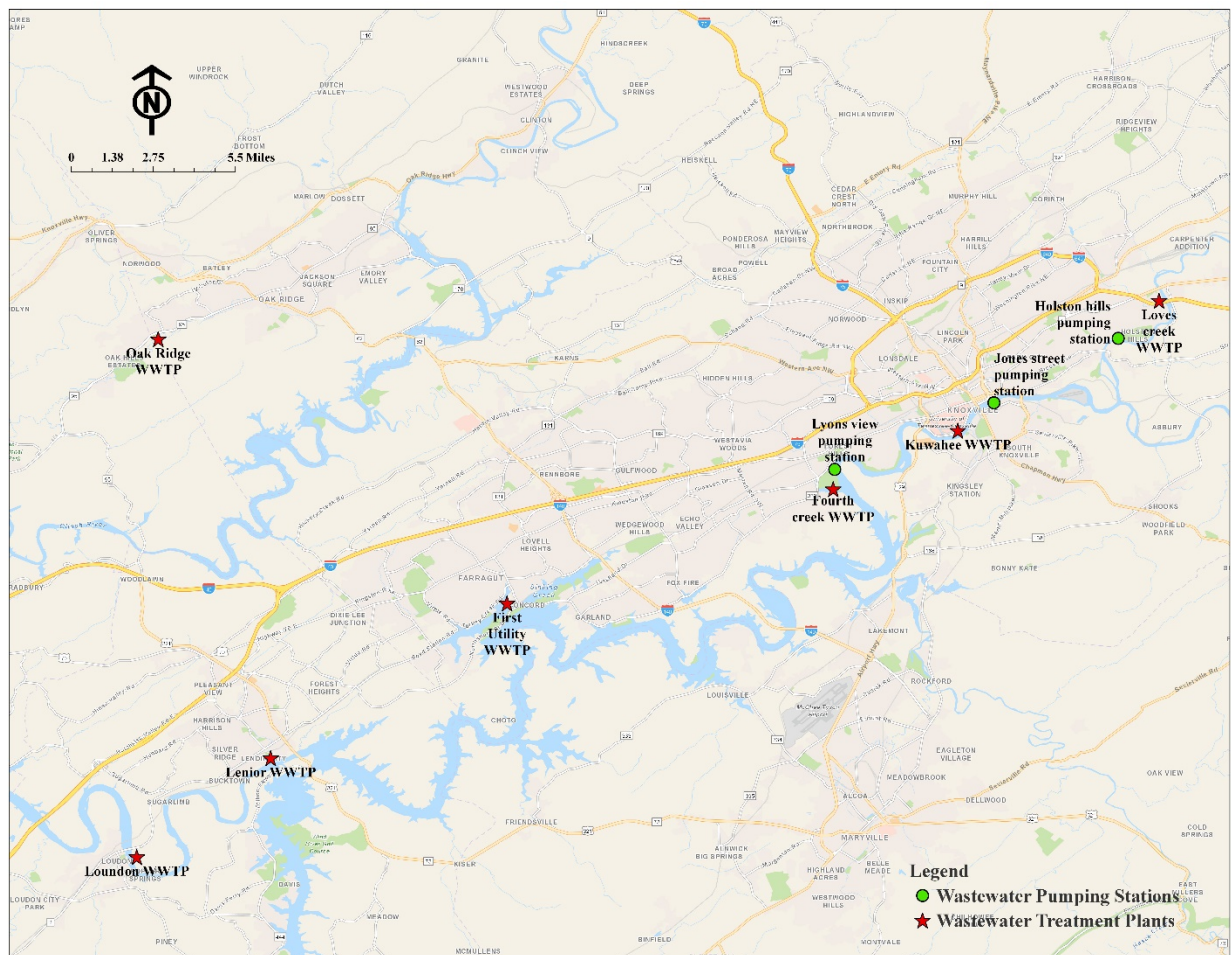


Figure 5-1 Sampling points selected in the studied wastewater treatment plants and wastewater pumping stations.

The red star labeled were wastewater treatment plants, while the green dots stand for wastewater pumping stations.

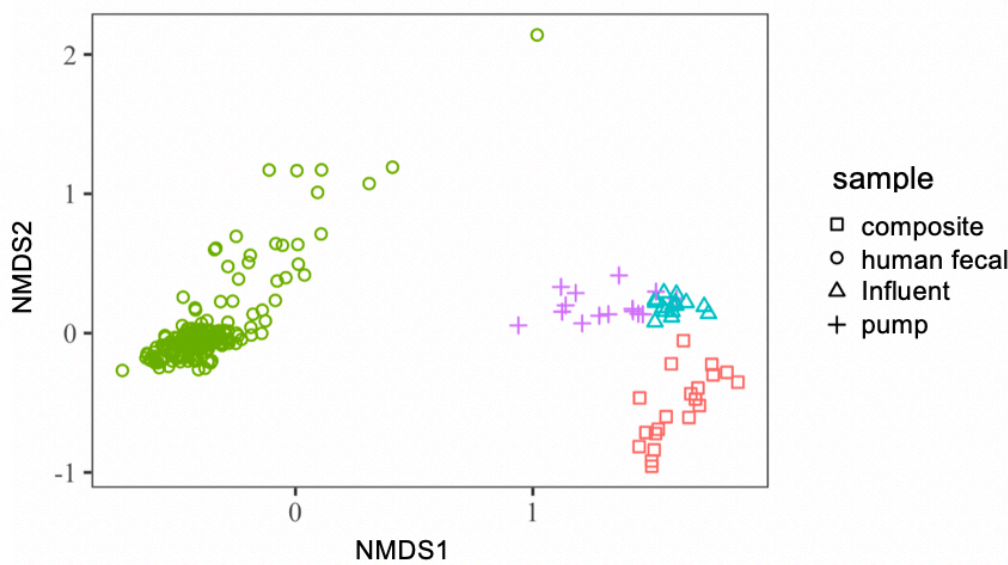


Figure 5-2 Sample relatedness of wastewater and human fecal microbiome at the genus level.

NMDS analyses were performed using unifrac dissimilarity matrix, which describes the relatedness of different types of wastewater and human fecal samples. Fecal stands for human fecal samples. Composite stands for flow-weighted composite samples taken over a 24-h period. ‘Influent’ stands for single-time-point grab samples from wastewater treatment plant influent. Pump stands for single-time-point grab samples from wastewater pump station in the residential area.

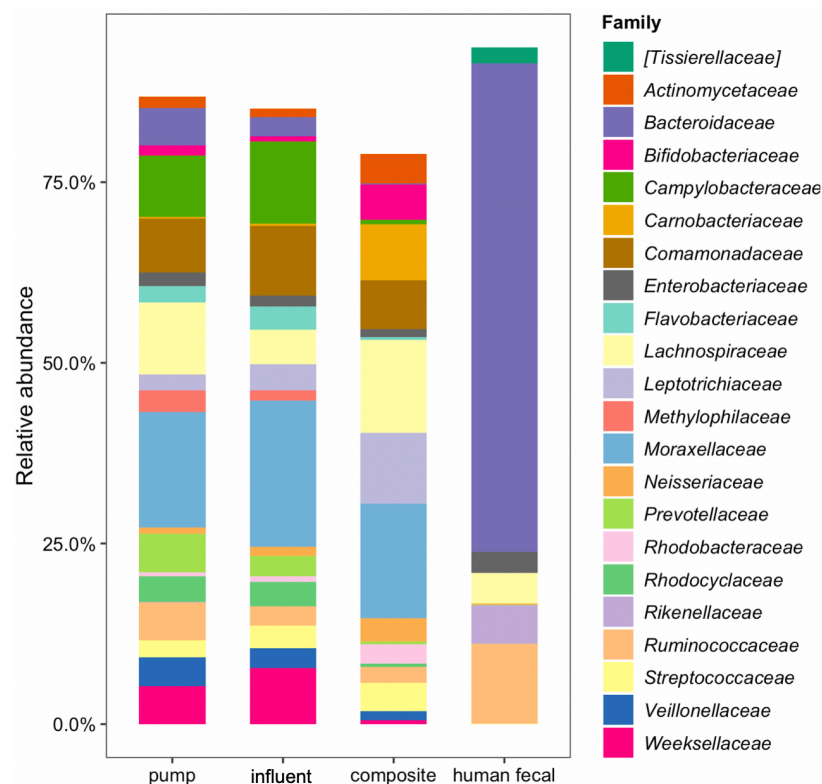


Figure 5-3 Taxonomy distribution of top abundant families in the human fecal samples and the pooled sewage samples.

The microbial community (relative abundance greater than 2%) in the sewage data set represents the total proportion of taxonomy composition from the top existed families in human fecal samples.

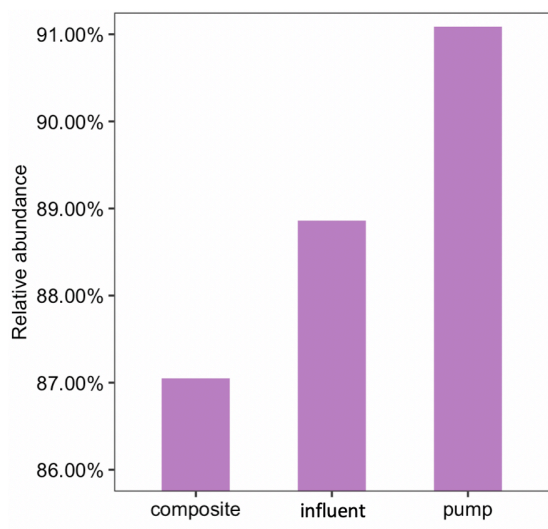


Figure 5-4 Taxonomy mapping ratio of all classified families in the human fecal samples and the pooled wastewater samples.

All families existed in human fecal samples with relative abundance existed in different types of wastewater.

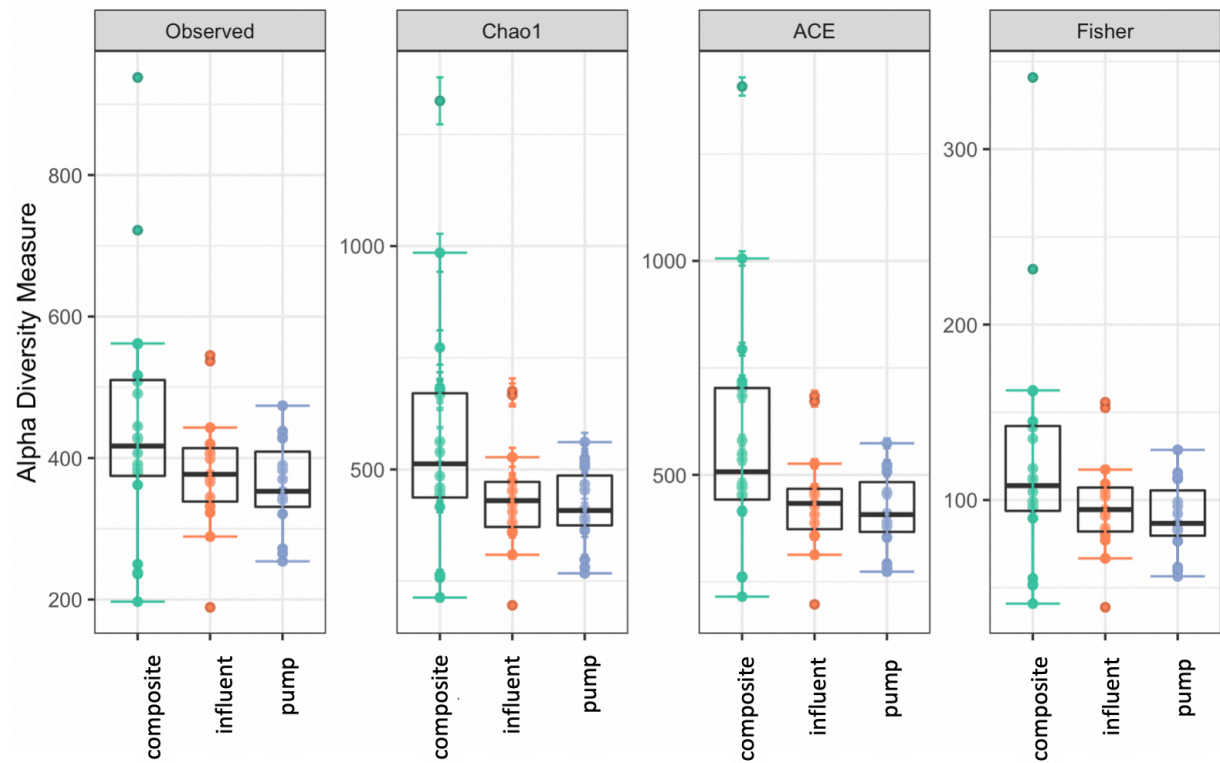


Figure 5-5 Alpha diversity measure of different types of wastewater samples.

Alpha diversity was measure with Observed OTU, Chao1, ACE and Fisher index.

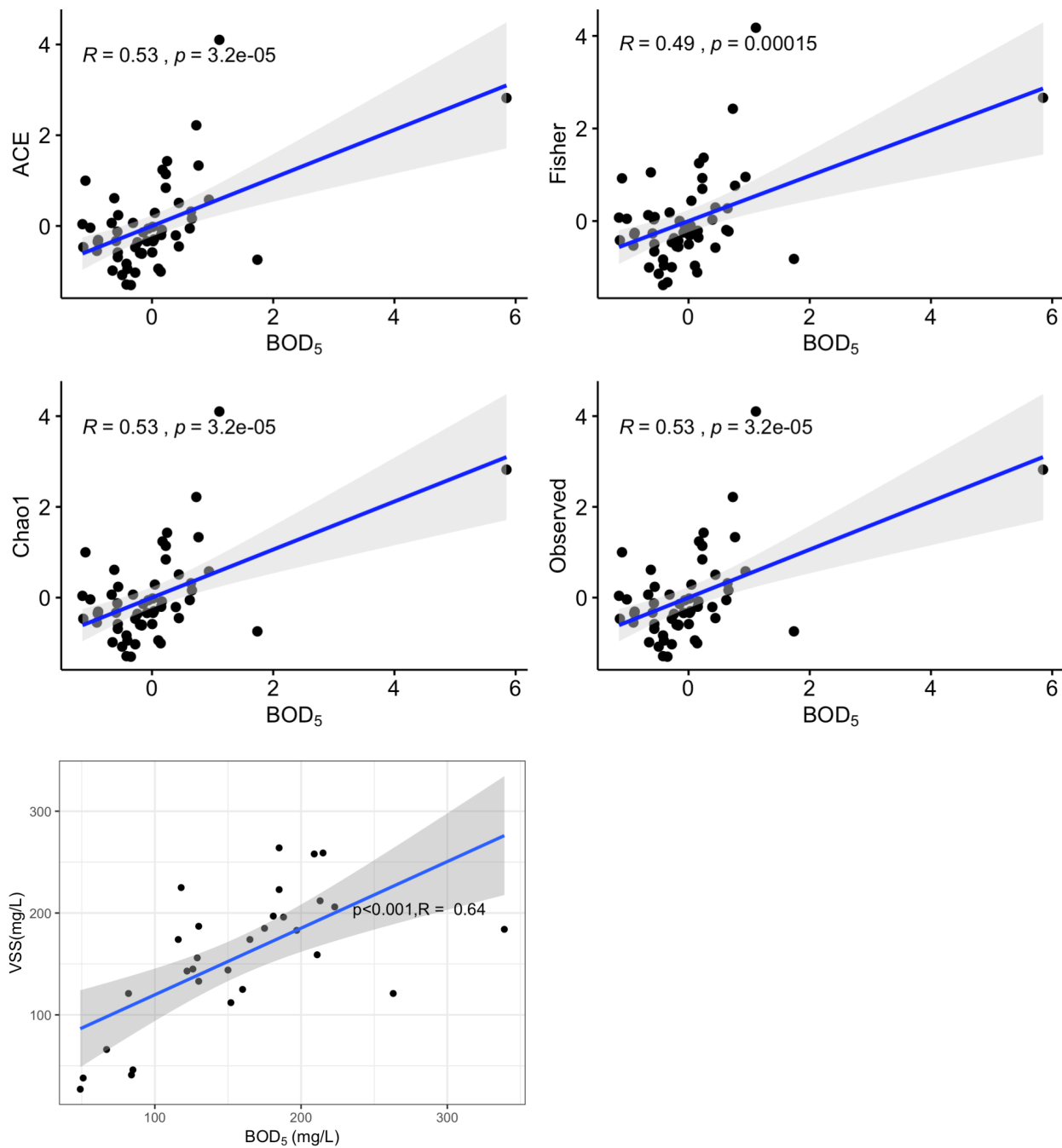


Figure 5-6 Relationship between BOD_5 and alpha diversity, VSS in wastewater samples.

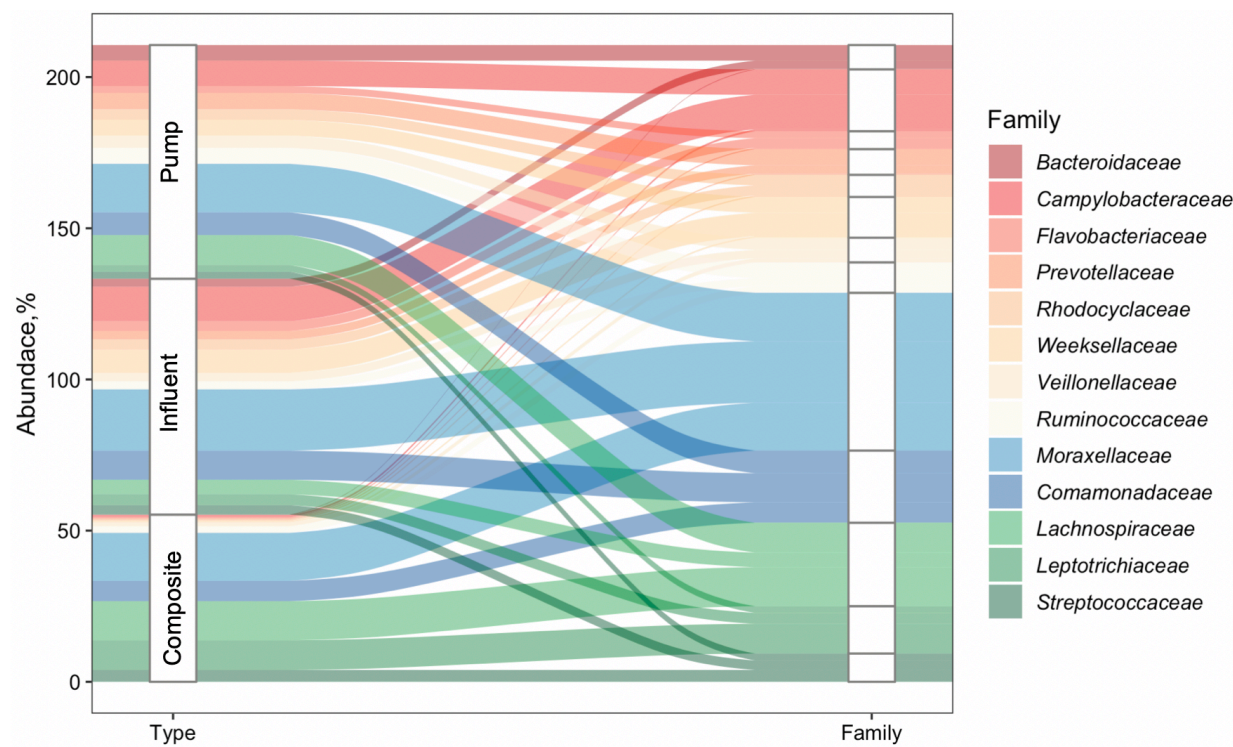


Figure 5-7 Comparison of pump station wastewater samples, influent wastewater samples and composite wastewater samples at the family level.

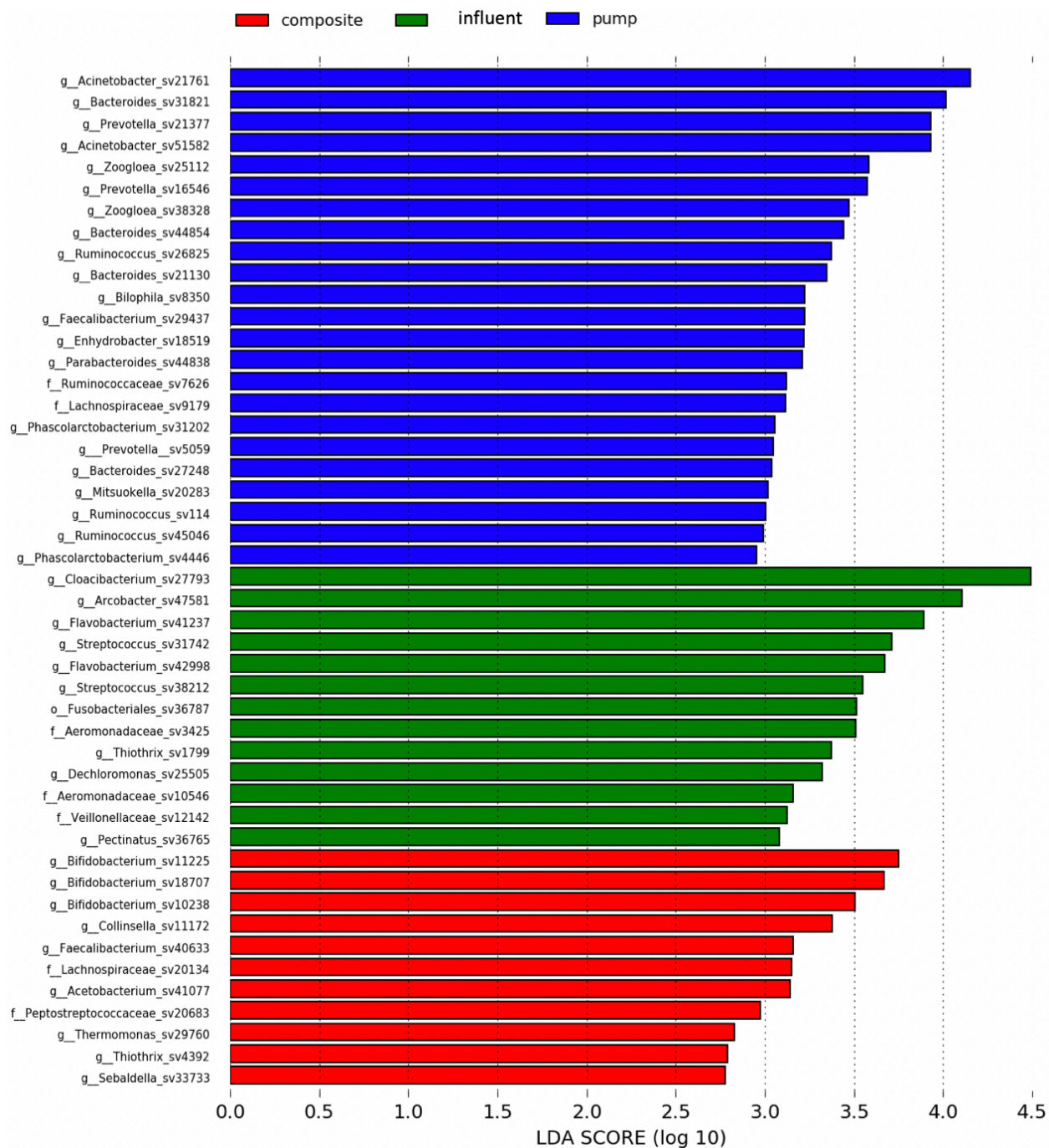


Figure 5-9 Biomarker detection analysis (LEfSe) on three types of sewage.

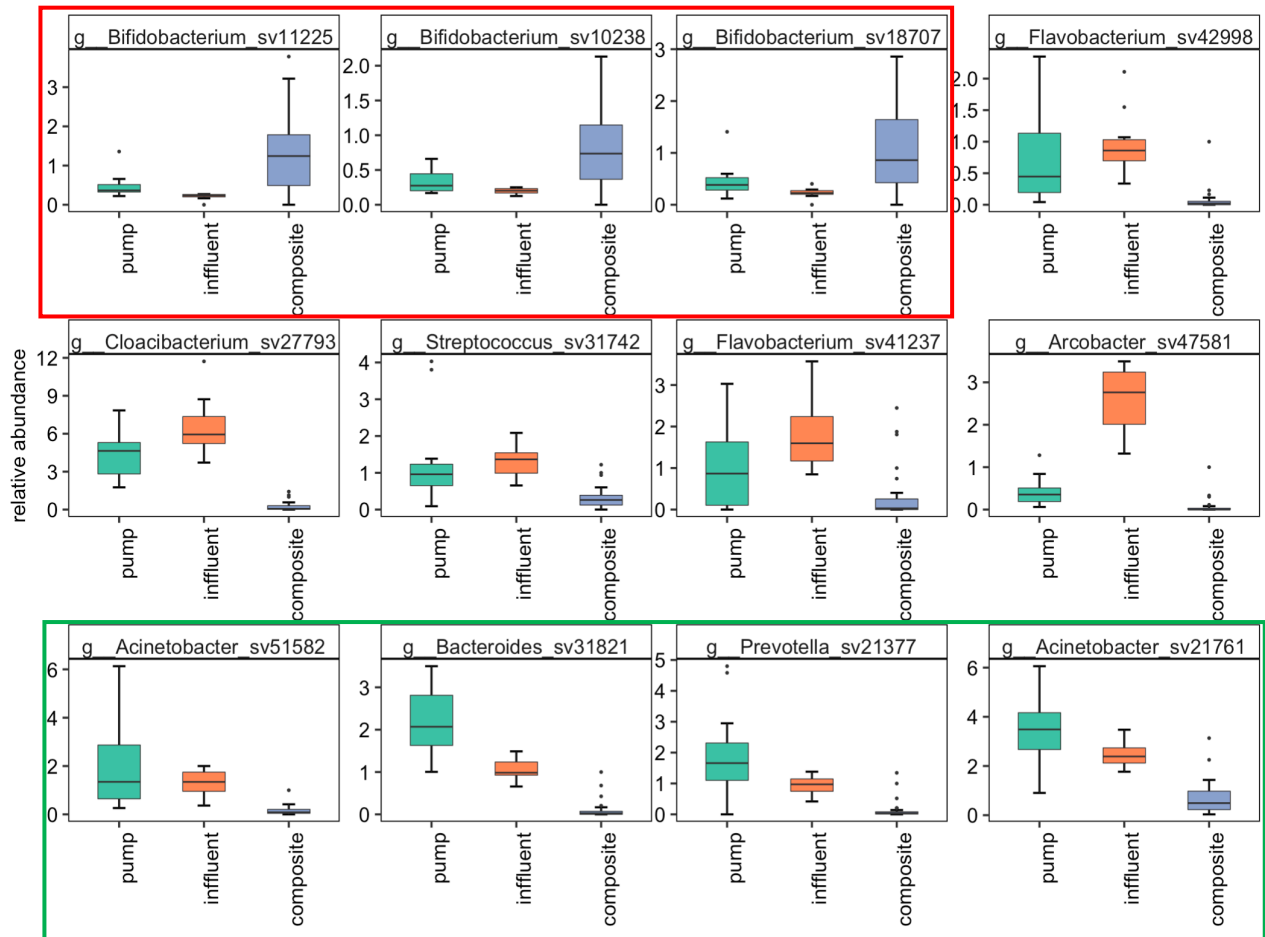


Figure 5-10 Results of top abundant bacteria found as biomarkers in each type of wastewater by Lefse biomarker discovery analysis.

The analysis was performed by first analyzed all sequence variants under Kruskal-Wallis test, while sequence variants violated the null hypothesis are further analyzed by the pairwise Wilcoxon test to check all pairwise comparisons between subclasses within different classes significantly agree with the class level trend. Finally, the resulting subset of vectors was used to build a Linear Discriminant Analysis model from which the relative difference among classes is used to rank the features. As a result, 23, 13, and 11 biomarkers were found in a pump station, wastewater treatment plant influent, and 24 hours composite samples. Boxplot only displays the top abundant ones. Red square part is biomarkers in composite samples, green square part covers pump station samples, while the rest of them are influent biomarkers.

(a)

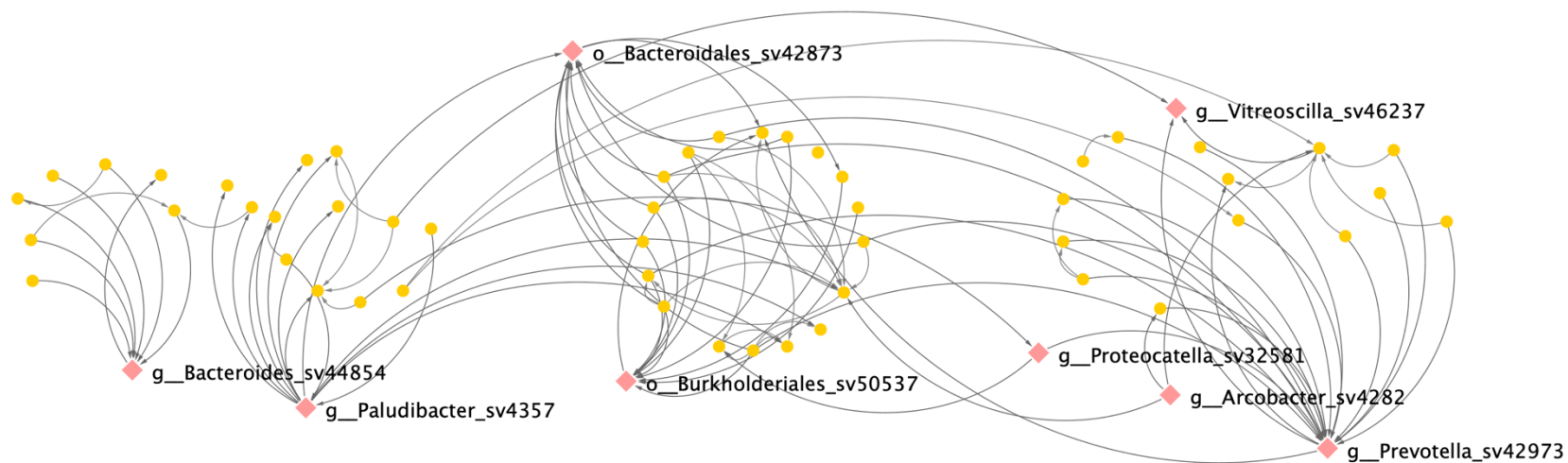


Figure 5-11 Subnetworks of each type of sewage.

a. pump station sewage; b. WWTP influent sewage; c. 24-hour composite sewage

The networks only displayed the network hubs, network module hubs and network connectors (squared nodes) with its first neighbours (yellow circle nodes).

(b)

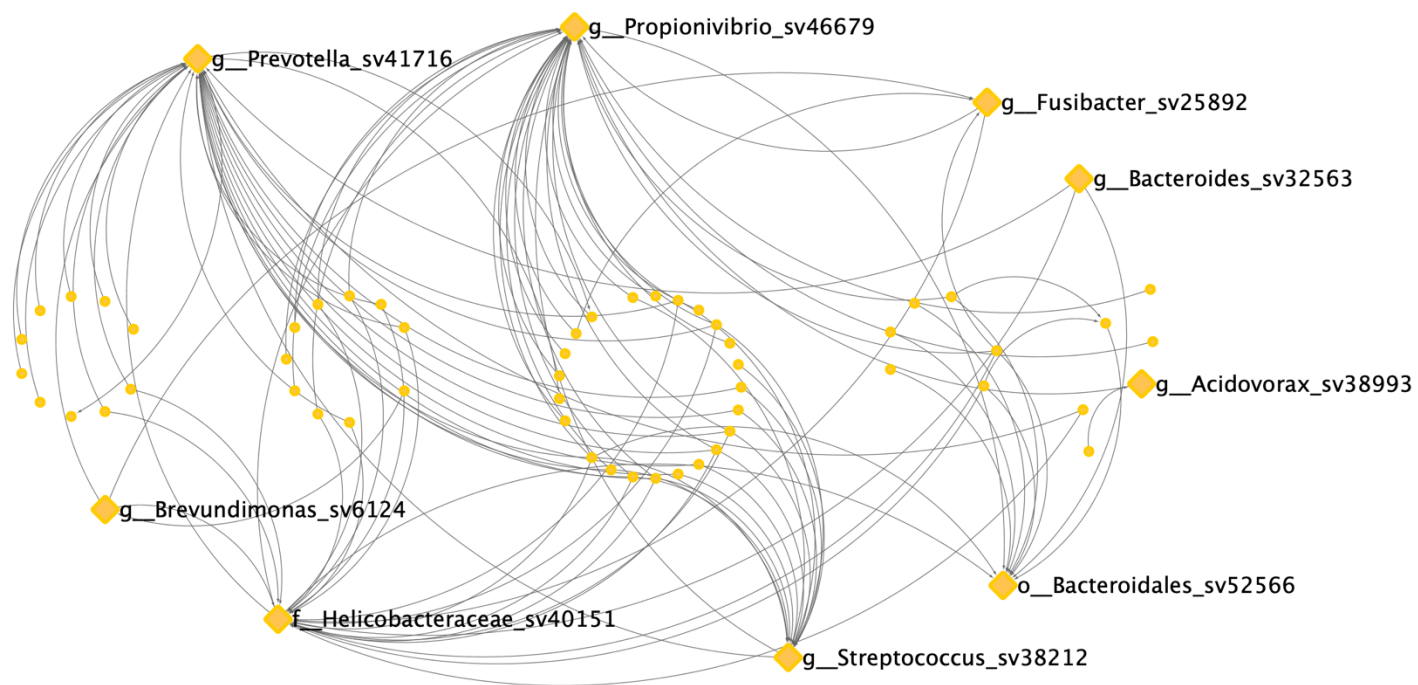


Figure 5-11 Continued

(c)

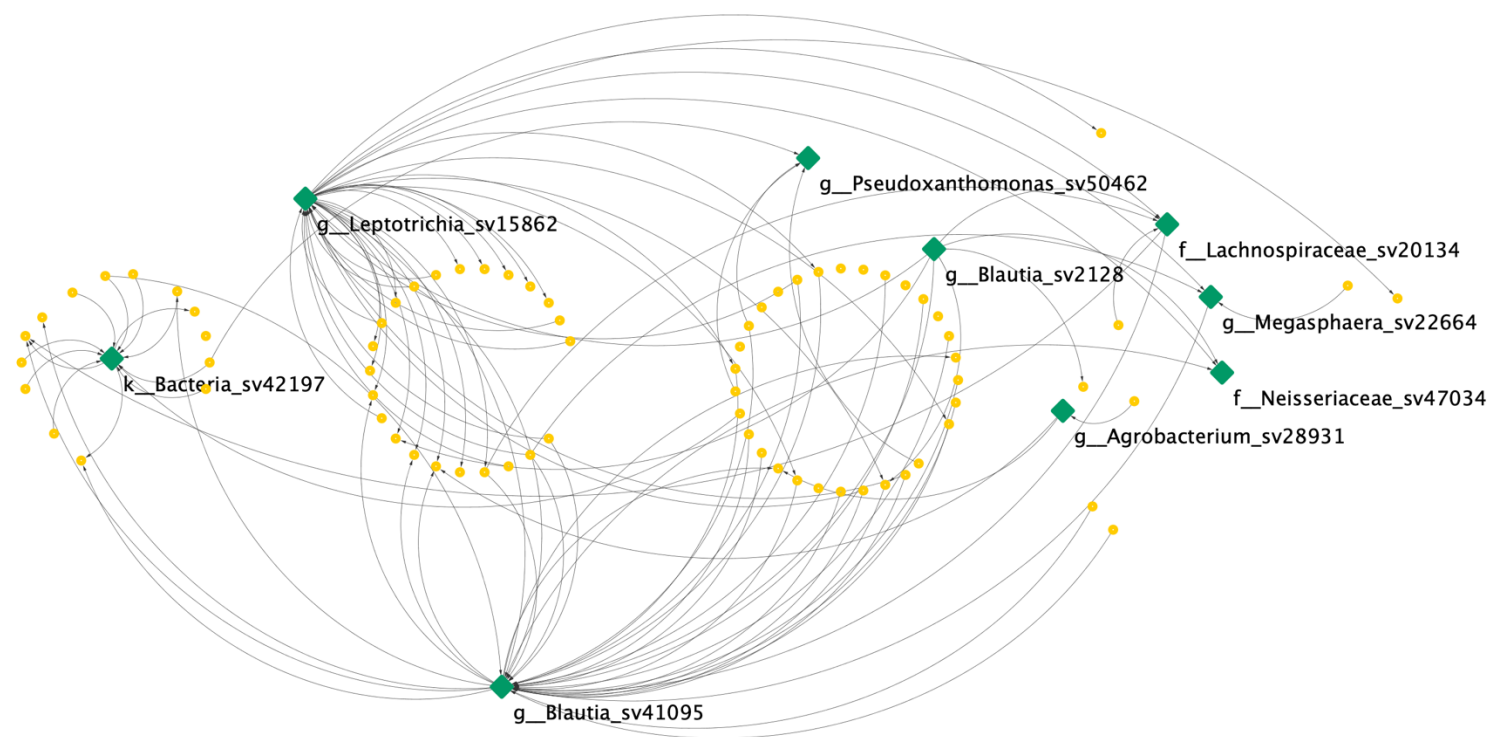


Figure 5-11 Continued

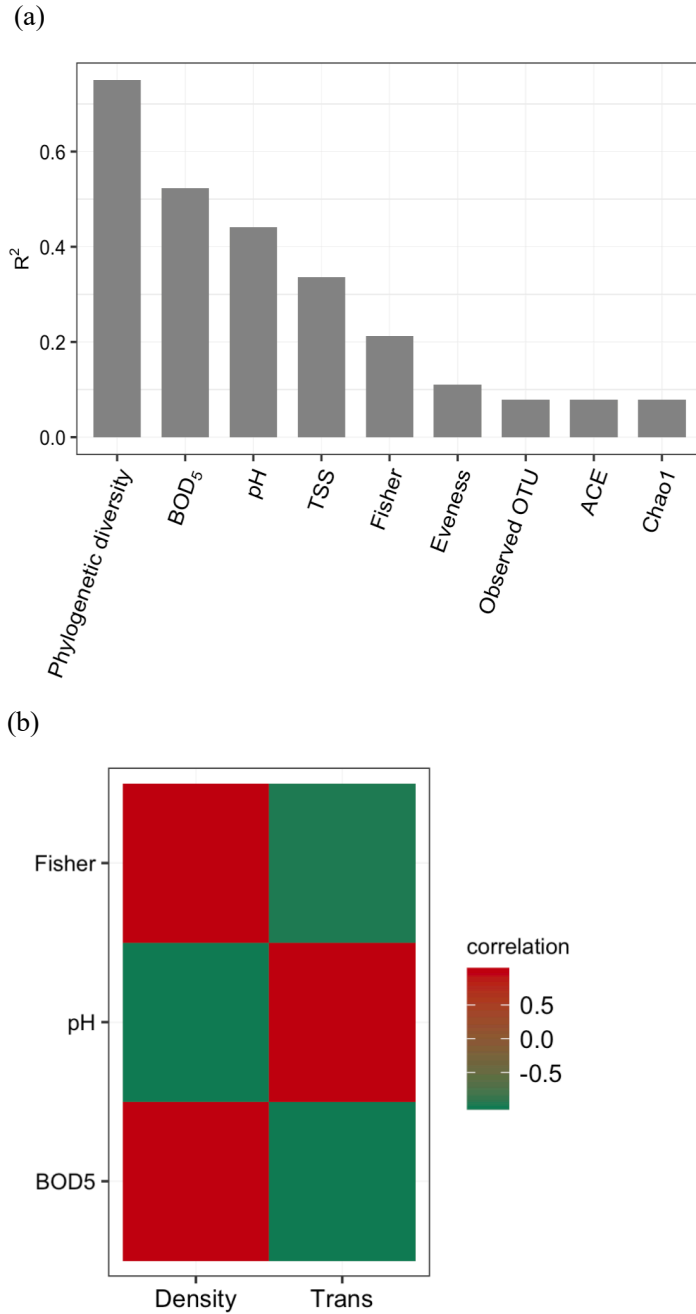


Figure 5-12 Relation between network characteristics and microbial composition, environmental variables.

(a) The contributions of variables related to network characteristics, (b) Pearson correlation between community composition or process variables and network features.

Section B: Stream Microbial Population in Response to Human Impact: A Perspective of Microbial Ecology

Abstract

The microbial community play essential roles in stream physiological function. Evaluating their responses to human disturbance is of great importance in ecological microbiology study. Here, we consider that the microbial changes in urban streams induced by human perturbations are pollution source oriented, which makes a healthy stream to an unhealthy state. The result verifies ‘Anna Karenina principle’ for stream water microbiomes, in which impacted stream water microbiome vary more than pristine streams—similar to Leo Tolstoy's dictum that “all happy families look alike; each unhappy family is unhappy in its own way.” The unhappiness in the stream is caused by different degree/type of human impact. The impact degree on stream water is assessed by microbial source tracking. Application of SourceTracker to stream water microbiome identified that sewage is a significant contaminant to stream baseline. The unhappiness in the stream baseline is identified by the prevalent of some sewage bacteria. The high concentration of *E.coli* and total coliform could arise from sewage contamination. MRM analysis reveals that the higher concentration of *E.coli* and total coliform relates to the lower microbial diversity. However, the main unhappiness in post-storm stream water by cause of stormwater runoff and sewage. Levels of coliform concentration in stormwater runoff reach far from the recreational water quality standards. A comparable level of *E.coli* and total coliforms in stormwater runoff and stream water signifies the integration of stormwater and stream water. In addition, land use could be an essential factor that influences the microbial community assemblage, which needs to be further studied. In contrast to unhappiness, stream happiness is embodied in the low variation of the microbial community in pristine streams. Lastly, even though human have profoundly influenced the stream ecological state, urban streams still reserve a group pristine microbiome consortium, which could be attributed by the resilience ability of the natural system.

Keywords: Anna Karenina; microbial ecology; source tracking; stream; sewage; stormwater

1. Introduction

Microbial populations play significant roles in stream ecosystems, which influence core processes in stream nutrient cycles (Findlay 2010). However, fecal contamination of surface waters is a continuous environmental problem which threatens ecosystems and human health risks in the United States (USEPA 2009). Clean Water Act Section 303(d) has been listed that an alarming amount of the United States rivers/streams (46%) are not safe for fishing and swimming according to the National Water Quality Inventory 2017 Report. Numerous sources contribute to fecal pollution, such as sewage, pavement runoff, animal feces, and others. Sewage leakage, caused by unhealthy urban plannings, has long been considered as the most critical point pollution (Lee et al. 2015, Pennino et al. 2016), which poses a severe risk to the health of the ecosystem. However, detection of leak sewers has always been a problem in urban water management, since sewage microbiome is of great variability (Mattioli et al. 2017), and only a few sewage markers are available to date (Dvory et al. 2018). In addition, pavement runoff is another critical cause of the degradation of the water environment (Petrucchi et al. 2014). It has been suggested that land-use, the impervious surface material types in transportation and construction, could influence the pavement runoff status (Fraga et al. 2016, Gnecco et al. 2005, Ibekwe et al. 2016). Therefore, the complexity of the source environment makes microbial source tracking even more urgent and complicated.

Source tracking has been discussed for a long time in assessing the fecal contamination of water contamination, such as *E.coli*, *Enterococcus* (Anderson et al. 2005). Traditional plate counting method can only distinguish fecal inputs or not, which is not available in source type identification. Then the qPCR method with different biomarkers, which is capable of identifying a variety of fecal sources, such as human, wastewater, animal-based markers (Kildare et al. 2007, Shanks et al. 2009, Yang et al. 2015). It has also been identified to be capable of figuring the contamination level of different sources. However, it has not been standardized (Mayer et al. 2016), and only a limited number of markers are available. Luckily, 16S rRNA amplicon sequencing method provides a new method, which comprehensively surveys the microbial populations with detailed information of amplicon sequence reads number. SourceTracker was developed to estimate the proportion of possible sources to the designated samples. It combines the Bayesian method in compliance with Gibbs sampling and Dirichlet prior parameters, which has shown high accuracy in environmental sample prediction (Henry et al. 2016, Liu et al. 2018, Staley et al. 2018). Therefore, application of this method in stream water study is of high confidence.

To date, most studies in community-based microbial source tracking focused on the accuracy improvement in the source prediction, engineered source tracking strategies. Stream plays an essential role in intense urban development for its abundant ecosystem services, consisting of water and erosion regulation, self-purification, nutrients cycling, water cycling, etc. However, few studies have considered microbial source tracking from microbial ecology aspects. It has been studied that microbial variation has been associated with an Anna Karenina principle (AKP) for microbiomes in animal/human-related research, which derives from the opening line of Tolstoy's Anna Karenina: “all happy families are all alike; each unhappy family is unhappy in its own way” (Zaneveld et al. 2017). It addresses the high similarity of microbiome composition under non-stress condition, while with high dissimilarity under stress control. It can not only help reveal patterns of microbial community change but also reflect microbial dynamics, which aids in understanding the host health from microbiotas. This theory has been applied to the animal/human microbial study, while little has been applied to environmental research.

In this study, we argue that AKP effects in stream water associated with stream water health. We apply the 16S rRNA technique to disentangle the variation and resilience of the microbial community in urban streams. Streams are dissected into happiness and unhappiness by AKP effects. We describe the unhappiness degree by microbial source tracking analysis. More specifically, we determine the source contribution to stream water under human impacts, evaluate the microbial indicators of different source, disclose the possible influencing factors of stream unhappiness. Our findings may provide a scientific background to stormwater management and urban infrastructure maintenance.

2. Methods

2.1 Sites description and sample collection

Study sites were selected from the urbanized area in east Tennessee, USA (Figure 6-1). All sample sites were near the city center downtown, which showed a high coverage of the developed area. These streams (stream A, stream B, and stream C) were classified as impaired streams in 303(d) list which was published by Tennessee Department of Environment and Conservation in 2016 (2016). The impaired reason was that microbial pollutant (*Escherichia coli*) exceeds water quality standards. The pollutants mainly came from municipal separate storm sewer systems (MS4s). We sampled stream water from sixteen storms events and their corresponding baseline. In addition, we also collected local pavement runoff samples around the watershed, local wastewater samples from the local wastewater treatment plants influents and pump stations. To exclude the stormwater influence to wastewater, 72-hour dry weather was ensured prior to sampling. For capturing a better microbial profile of residential sewage, all sewage samples were sampled at 8 am in the morning. Moreover, some possible animal microbial sequences were downloaded from the European Nucleotide Archive database.

Stream water and pavement runoff (~ 4L) were collected in sterile, autoclavable, 2-liter Nalgene containers, which were pre-sterilized with a light bleach solution and rinsed thoroughly with distilled water. Pavement samples were collected just right after the rain stopped. All samples were transported at 4 °C (on ice) and stored at -20 °C until DNA extraction. Coliform bacteria were tested within 6 hours. *Total Coliform* bacteria were determined by traditional membrane filtration techniques (APHA, 2005). *E. coli* were identified by the EPA-approved membrane filtration method using m-ColiBlue24® liquid media (HACH, Loveland, Colorado, USA).

2.2 DNA extraction and DNA sequencing

DNA extractions were performed using MP FastDNATM Spin Kit for Soil according to the manufacturer's instruction (Protocol Revision #116560200-201608). Briefly, the biomass samples were resuspended in the 978 µl Sodium Phosphate Buffer solution and 122 µl MT Buffer, then followed by 100 seconds vortex on MP biomedical fast prep instrument. Then the DNA was purified with Zymo Genomic DNA Clean & ConcentratorTM-10 kit according to the instructions provided by the manufacturer. The DNA concentrations were determined using Thermo Scientific NanoDrop ND-3300 Fluorospectrometer.

Polymerase chain reactions were prepared with a cocktail mix containing 12.5 µL Phusion flash Master Mix, 10 µL ultra-pure water, 1 µL forward primer, 1 µL reverse primer, and 2 µL (100 to 150 ng) DNA template. Each universal primer pairs were designed to target the V4 region of bacterial and archaeal 16S rRNA genes: forward primer 515F (GTGCCAGCMGCCGCGGTAA) with 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACAC), forward primer pad (TATGGTAATT), and forward primer link (GT); reverse primer 806R (GGACTACHVGGGTWTCTAAT), with adapter (CAAGCAGAAGACGGCATACGAGAT), link (CC), and a unique 12-base specific barcode for each sequence. The PCR program included one cycle of 94 °C for 3 minutes, followed by 35 cycles at 94 °C for 45 seconds (denaturation), 55 °C for 1 minute (annealing), and 72 °C for 1 minute and a half (elongation), a final extension at 72 °C for 10 minutes, and then stored at 4 °C. After the program finished, the samples were run on an Agilent 2100 Bioanalyzer Instruments with Agilent DNA 7500 chips to ensure the amplicon quality and measure the DNA concentrations. Post PCR, amplicons were pooled (a set of 9 or 10) based on the peak height from Agilent DNA 7500 analyses and the concentrations were measured using KAPA Illumina kit for quantifying library (KK4824) from KAPA Biosystems. Samples were pooled again according to the results and number of sequences desired into a final one and the concentration was measured again with the KAPA quantifying kit. 4 nM pooled libraries were diluted with 10 mM Tris/0.05% tween buffer (pH = 8.5) and denatured with 0.2 N freshly made NaOH solution. The sample was then combined with 20% PhiX control kit to increase the library diversity. The mixture was incubated at 96 °C for 2 minutes using a heat block and then put into a water-ice bath for 5 minutes. After the sample was loaded, an Illumina MiSeq System with a MiSeq Reagent Kit v2 (300-cycles) and Metagenomic workflow was selected to execute the 16S protocol using the MiSeq Reporter software (MSR) at the University of Tennessee Genomics Core (UTGC).

2.3 Bioinformatics analysis and statistical analysis

All sequencing data obtained from Illumina MiSeq platform runs were denoised under QIIME2 pipeline DADA2 algorithm. For matching the sequence length obtained from the European Nucleotide Archive database collected data, sequences downloaded were sequenced with the same primer and similar library setup procedures. A mapping of feature sequences was summarized, and a pre-trained Naive Bayes classifier based on Greengene 13_8 99% OTUs were used in the taxonomy analysis. All the statistical analysis was conducted in R (Team 2016). Principal coordinate analysis (PCoA) and alpha diversity estimation were performed in *phyloseq* package, principal component analysis with biplot was performed in *factoextra* package. Log2Fold change of microbiome was analyzed in *DESeq2* package. Tukey HSD test and correlation analysis were implemented in base R. Analysis of MOlecular VAriance (AMOVA) was carried out in *poppr* package. All plots were plotted under *ggplot2* package. Source prediction was predicted in SourceTracker2. Multiple Regression on distance Matrices (MRM) analysis was performed in *ecodist* and *vegan* package. The phylogenetic tree was constructed in MEGA. Hydrology map was downloaded from Tennessee GIS Data & Resources website (<https://libguides.utk.edu/tngis>). Land cover map 2011 was downloaded from USGS website (<https://nationalmap.gov/landcover.html>). The map was plotted, and the watershed area was calculated in ArcGIS version 10.6.

3. Results

3.1 Bacterial community structure among samples

Water samples were collected from three different urban streams in east Tennessee, USA. The V4 region of the 16S rRNA gene was sequenced from stream water. Considering from the generalized phylum level, urban stream water samples were mainly accounted by *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Verrucomicrobia*, and *Cyanobacteria*, and *Bacteroidetes* (Figure 6-2), which was similar with pavement runoff and reference stream microbiome. This result was similar as previously listed urbanized streams and storms (Baral et al. 2018, Fisher et al. 2015, Huang et al. 2018, Sun et al. 2016). However, Animal feces show the distinct different proportion of microbiome, the relative abundance of *Firmicutes* and *Bacteroidetes* account up more than 80% of the community. Sewage is in between the natural water and animal feces, with higher relative abundance in *Firmicutes* than natural water, and higher abundance in *Proteobacteria* than animal feces. Besides, it should be noted that stream B showed the higher relative abundance of *Firmicutes* and *Bacteroidetes* than creek A and creek C. As previously described, some genus in *Firmicutes* were identified as fecal sources from sewage (McLellan et al. 2010c, Shanks et al. 2013). It may imply that stream B could be influenced by sewage. Based on this assumption, we further take a closer comparison at the top abundant sequence variants in stream water and fecal sources.

Top abundant existed sequence variants existed in animal feces and sewage and also existed in urban streams greater than 0.1% were displayed in Figure 6-3. These sequence variants belong to *Proteobacteria*, *Fusobacteria*, *Firmicutes*, and *Bacteroidetes*. Interestingly, conform to former guess that stream A and stream C got a lower abundance of fecal related microbial than stream B. Only a few dots were shown in stream A and C, but much was filled in stream B. Furthermore, even in creek B, dry day samples and showed the more abundant microbial community in *Bacteroides* and *Prevotella*. Both bacteria have been used as biomarkers in detecting human and ruminant fecal pollution (Bernhard and Field 2000). However, they are easy to decay in natural water (Bae and Wuertz 2009). Therefore, the higher abundance of these bacteria in dry day samples than wet day samples could be attributed to the newly fecal source inputs. These fecal inputs could be flushed away by the storms during wet days. Moreover, a group of other bacteria also existed in urban streams, such as sequence variants from *Arcobacter*, *Acinetobacter*, *Acidovorax*, *Methylophilaceae*, *Enterobacteriaceae*, *Leptotrichia*, *Megamonas*, *Faecalibacterium*, *Peptostreptococcaceae*, and *Cloacibacterium*. Most of them were found as significant existed bacteria in sewage. It may indicate the high probability of sewage contamination in urban streams.

Principal coordinates analysis with the Unifrac distance calculates the microbial community relatedness among possible source samples and stream water samples. It aims at checking the dissimilarity among samples (Figure 6-3). PCoA showed that stream water samples and sewage/animal feces samples clustered separately indicated the microbiome in sewage/animal feces is different from that found in natural water. AMOVA pairwise comparison identified significantly different microbial community composition among sewage, animal fecal samples, stormwater runoff samples, and stream water samples. Fecal samples from the same source type located in the same color, sewage clustered in a single color. However, some urban stream samples showed more relatedness with sewage samples, while some were closer to stormwater runoff samples. It may imply that some urban water samples suffer from sewage contamination, while others sustain stormwater runoff flushing.

3.2 Human stress induces microbial AKP effects in streams

For determining possible sources contribution, Bayesian-based classification software SourceTracker2 was applied in the estimation (Knights et al. 2011). Five independent runs were performed on the highest sequence variants resolution under the default settings (Figure 6-5). It predicted that sewage is likely the most common input source in urban streams baseline, while dog and calf source are found as well. As expected, stream B bear a higher proportion of sewage inputs than stream A and stream C. 5/20, 6/20 samples were detected with calf and dog feces pollution. However, the likely most common source inputs change to stormwater runoff, followed by sewage, and bird feces. Only 2/20, 1/20, 2/20 samples respectively found bird feces, calf feces, and dog feces pollution, with a proportion higher than 1%.

To identify the determinant factors of the microbial community composition, a principal component analysis was carried out with environmental factors/sources contribution and urban stream microbial community (Figure 6-6). Baseline samples and post-storm samples separated. Baseline samples related to sewage, dog feces, and calf feces. Some post-storm samples related to stormwater runoff source, some related to precipitation, total coliform, and *E.coli*. Additionally, ADP may have a relation with *E.coli* and precipitation.

A MRM analysis revealed that coliforms have a negative relation with the microbial community ($R = -0.84$, $p = 0.04$), but the microbial community has a significant positive relation with phylogenetic diversity ($R = 0.94$, $p = 0.02$) (Table 6-1). It indicates that the increased coliform bacteria may suggest the decreased phylogenetic diversity, which means the die off or decay of some bacteria with low survival ability. Under dry weather condition, increased coliform bacteria could be associated with sewage inputs. Significant positive relationship was found between sewage contribution proportion and coliforms, *E.coli* ($R = 0.76$, $p < 0.05$), total coliform ($R = 0.63$, $p < 0.05$) (Figure 6-7). Under wet weather condition, increased concentration of *E.coli* in stream showed significant positive relation with survived *E.coli* in stormwater runoff (Figure 6-8). It may enhance the conclusion that *E.coli* and total coliform mainly attributed to sewage and stormwater runoff respectively in dry weather and wet weather.

However, coliform concentration in baseline was significantly lower than stormwater runoff and post-storm stream water (Figure 6-9). Based on our experiment test of *E.coli* concentration in sewage, sewage sustained the concentration of *E.coli* around 5 million CFU/100mL and total coliform around 16 million CFU/100mL. It is much higher than the *E.coli* concentration in stormwater runoff, which is around the mean value of 18700 CFU/100mL. Stream water was mainly polluted by sewage in dry weather and stormwater runoff in wet weather. Therefore, baseline samples would be in a higher concentration of coliform than post-storm samples. However, baseline samples were significantly lower than the post-storm samples. Such an interesting finding could be due to the strong ability of sediment. Sediment can provide a more friendly place for *E.coli* survive (Burton et al. 1987, Laliberte and Grimes 1982), regrow (Litton et al. 2010) and accumulate (Alm et al. 2003). Sewage leakage is a chronic problem, which may support sediment accumulate sewage-derived microbial source. The activation of sediment in the river/stream bed could occur when critical shear stress threshold of the stream sediments was exceeded (Bradshaw et al. 2016, Cho et al. 2010), and the resuspended fecal indicator bacteria could propagate in a very long distance in the stream. Therefore, the survival ability of fecal indicator bacteria makes sediment an important sources during the storm events (Cho et al. 2010, Garzio-Hadzick et al. 2010, Yakirevich et al. 2013). And this could contribute to increased *E.coli* concentration in the post-storm stream water.

3.3 Unhappiness in AKP effects cause decreased microbial diversity

Reference streams showed the highest alpha diversity, followed by baseline, then post-storm stream samples. This pattern was in reverse relation to coliform concentration. The lower concentration in coliforms, especially *E.coli*, the higher microbial diversity (Figure 6-10). It may suggest that the higher concentration of coliforms, the higher pollution sources contribution, and resulted in the lower microbial diversity. The inputs of sewage and stormwater runoff cause the primitive ecological microbiome to die off and decay. Some accompanying microbial immigrants, originate from human impact, colonize the urban stream system, and reshape a new stream microbial consortium. How the microbial change under human impact was depicted by DESeq analysis. Consistent with source tracking result, urban streams express the invasion of sewage-related bacteria, such as *Prevotella*, *Blautia*, *Arcobacter*, *Acinetobacter*, *Flavobacterium*, *Alkanindiges*, *Flavobacterium*. Besides, some soil and aquatic bacteria also existed in urban streams, such as root colonized bacteria *Massilia*, *Pedobacterm*, *Flectobacillus*, *Dechloromonas* can stay in aquatic sediment environment. However, a group of bacteria that hard to classified in family or genus level existed in the reference stream. The low classification means that low discovery of these bacteria in the study to date.

Reference streams show more stable microbial composition than urban streams. PCoA plot depicts the close distance among reference streams and sparse distance among urban streams (Figure 6-11). The closer distance reflects more close microbial similarity. Therefore, reference streams even with more diversity still showed more stable microbial change, while urban streams subjected to more dissimilarity. It could be stream unhappiness,

human impacts caused the dissimilarity. Log2Fold change analysis was performed to discover the difference between two types of the stream. A group of the sewage-related microbiome was found in urban streams, which further revealed the unhappiness caused by human (Figure 6-12). On the contrary, reference streams were covered by the pristine microbiome, related to soil and water.

Even urban streams with unhappiness, a group of the microbial community still has its ability to recover from the environmental disturbance and reshape the niche of urban streams. Stormwater runoff flushing is a typical case. Stormwater showed more similarity with the post-storm sample. Most baseline samples showed more distance with stormwater, which may signify the recoverability of streams from perturbations. More details can be found in Figure 6-13, stormwater runoff showed more similarity with post-storm than baseline samples at the family level. Stormwater runoff showed higher abundance in *Oxalobacteraceae*, *Comamonadaceae*, *Moraxellaceae*, *Pseudomonadaceae*; they are a typical family in stormwater runoff. However, these bacteria decreased after a period of dry weather. Moreover, baseline samples showed the decrease of those storm condition increased families, while more families with low taxonomy classification increased. Therefore, stream water has its ability to recover from human stormwater disturbance. On the other hand, it may signify that streams without human impact preserve more natural microbiome, and these natural microbiomes can preserve high composition stability. Furthermore, the family level comparison revealed that stream under human perturbation or not still has a similar microbial community (Figure 6-13), such as *Comamonadaceae*, *Cytophagaceae*, *Flavobacteriaceae*, *Sphingomonadaceae*, *Rhodobacteraceae*. Most of them associate with soil, aquatic, freshwater environment. However, the streams in this study suffer from frequent sewage leakage, which prevents its long-term resilience process. Therefore, stormwater management and urban infrastructure maintenance become more significant.

3.4 Potential influence of land use cause the AKP effects

Three studied urban streams showed the different contribution of sewage inputs; stream B was much higher than stream A and stream C (Figure 6-14a). Correspondingly, stream B was on the highest level of nutrients, NO_3^{2-} , Cl^- and SO_4^{2-} (Figure 6-14b). Stream B showed a significantly different level of sewage contribution and nutrients concentration, revealed by Tukey HSD test ($p < 0.05$). Therefore, the unhappiness from sewage leakage caused the increase of nutrients in streams.

Since all three sites were urbanized areas that have been developed for a long time. For identifying the developing status of the different site. Developed areas along each watershed were calculated based on land cover data. Based on USGS land cover classification standards, the developed area includes open space area, low intensity developed area, medium intensity developed area, and high intensity developed area (Figure 6-1). The absolute area cannot represent the developing level for comparison. Therefore all area was standardized by each class area divided by the total developed area. As a result, stream B was much more developed than the other two streams (Table 6-2). The high intensity and medium intensity developed area have a higher percentage than the other two streams. Stream B had a total of 48.5% of high and medium developed area percentage, while stream A and stream B only had 24.6% and 30.6% of high and medium developed area. This may also suggest human impact intensity has impacted the urban water body, which may also imply why the stream B always showed much higher fecal source levels. This result is similar as (Bradshaw et al. 2016) listed that land use played a clear role in pathogen occurrence.

4. Discussion and Conclusion

Water quality in the streams, especially urban streams, has long been a discussing point. The microbial community analysis provides a new perspective on water quality evaluation. Evaluating stream water microbiome response to human impact by community based microbial source tracking verifies ‘Anna Karenina principle’ in environmental studies. Results support that the unhappiness in the stream is caused by different degree/type of human impact, mainly expressed in sewage and stormwater runoff. The microbial population changes caused by human perturbation would influence ecosystem processes. Detailed conclusions include:

The unhappiness in the stream water baseline was identified by the prevalent of some sewage bacteria. The high concentration of *E.coli* and total coliform could arise from sewage contamination. MRM analysis revealed that the higher concentration of *E.coli* and total coliform relates to the lower microbial diversity.

The main unhappiness in post-storm stream water mainly originated from stormwater runoff and sewage. Levels of coliform concentration in stormwater runoff reach far from the recreational water quality standards. A comparable level of *E.coli* and total coliforms in stormwater runoff and stream water signified the integration of stormwater and stream water.

Land use could be an essential factor that influences the microbial community assemblage. Stream B suffered more from the sewage leakage than the other two streams, which expressed by the existence of some sewage-originated bacteria. Along stream B, 17.2% of the land was under high developed status, while the other two streams only show less than 10% of the highly developed land use type. Land use is supposed to induce microbial changes which influence stream ecosystem function (Gucker et al. 2009). Because land use can change the total mass and lipid composition and affect the trophic transfer of energy in stream food webs, then influences the stream microbial food web (Boechat et al. 2011). However, a more comprehensive study needed to be implemented to identify this hypothesis.

Stream happiness was reflected in the low variation of the microbial community in pristine streams. PCoA identified closer microbial distance among reference stream samples than among urban streams.

Even though human have profoundly influenced the stream ecological state, urban streams still reserve a group pristine microbiome consortium. It could be attributed to the microbial resistance, while a group of the microbial population remains unchanged in the human disturbance. It suggests that those microbiomes are sensitive to environmental stress. Except for resistance, microbial resilience was also found in the stream water, while stream water microbiome was able to return to its original state. Among the thirty-four collected samples, fourteen of them were baseline and post-storm pairs. However, the microbial community in baseline water samples showed more similarity and post-storm water samples was in another group. It may suggest that the storm impacted the stream water, while stream water was able to bounce back after a period of recovery. How the water microbial recovered remains to be studied.

Community-based microbial source tracking with SourceTracker is library dependent, while a large proportion of sources is unpredictable. Prediction coverage is built on the library de novo established. In this study, a large number of sources remains unknown, which suggests the complexity of pollution sources is beyond our expectation. Sediment, especially the water-sediment contact surface, could be an ideal reservoir bacteria, such as fecal bacteria, waterborne pathogens (Alm et al. 2003, Bradshaw et al. 2016, Cho et al. 2010). However, during the storm event, re-entrainment of bed sediments resuspended since the critical shear stress threshold of the stream bottom sediments is exceeded (Cho et al. 2010, Jamieson et al. 2005). Therefore, including sediment as a vital source in a further study is hypothetically required.

Recently, negative control has been proposed in source tracking studies. It is of great importance in increasing prediction accuracy. For examples, wastewater effluents have high similarity with ocean water in the microbial community. When (Staley et al. 2017) predicted the pollution source contributed to endangered coral reefs, decreased false positive estimation will be if background ocean water was included in the source library. This

argument was further identified by incorporating either the local background or representative bacterial composition in water resources (Hagglund et al. 2018). And the measured coliform concentration was correlated with source tracking result. Therefore, we include reference water as background water in source prediction. Similar to previous results, we found a significant correlation with coliform concentration and predicted source contribution. Furthermore, if we emit the reference stream water from background library, we saw a significant exaggeration of stormwater runoff influence, with an average increasing stormwater runoff contribution of more than 5%. Besides, we also see a significant stormwater runoff influence in the dry period. However, this impractical finding disappeared if we included reference stream water in the library. Therefore, our study evidence the former hypothesis in background water inclusion.

Appendix F: Figures and Tables

Table 6-1 Results of MRM analysis to disclose the microbial community-related factors.

	microbial community	p-value
Intercept	9.71	0.34
Coliforms	-0.84	0.04
Sources contribution	0.68	0.10
Alpha diversity	-0.22	0.50
Phylogenetic diversity	0.94	0.02

Table 6-2 Different levels of the developed area along the studying streams.

	stream A(mi ²)	Percentage	stream B(mi ²)	Percentage	stream C(mi ²)	Percentage
Developed, Open Space	6.52	40.7%	1.22	19.0%	5.25	32.4%
Developed, Low Intensity	5.56	34.7%	2.09	32.6%	6.00	37.0%
Developed, Medium Intensity	2.86	17.8%	2.01	31.3%	3.36	20.8%
Developed, High Intensity	1.08	6.8%	1.10	17.2%	1.59	9.8%
Total Developed	16.02	100.0%	6.42	100.0%	16.19	100.0%

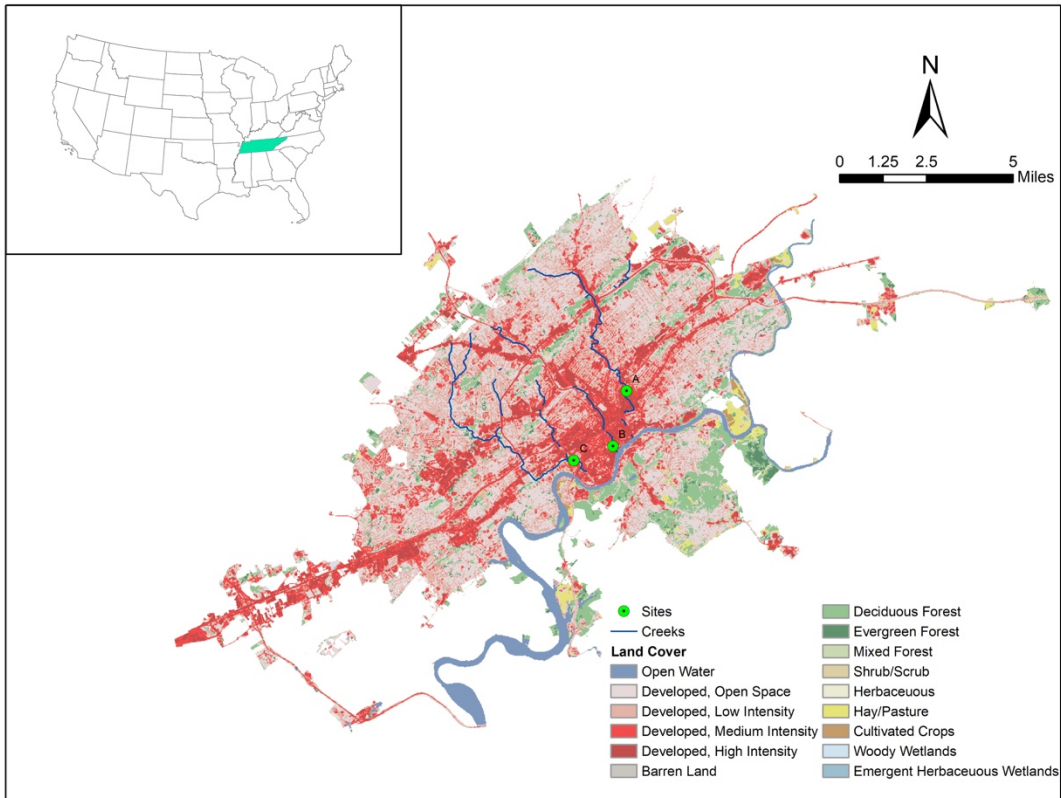


Figure 6-1 Study sites and the land cover of the developed area.

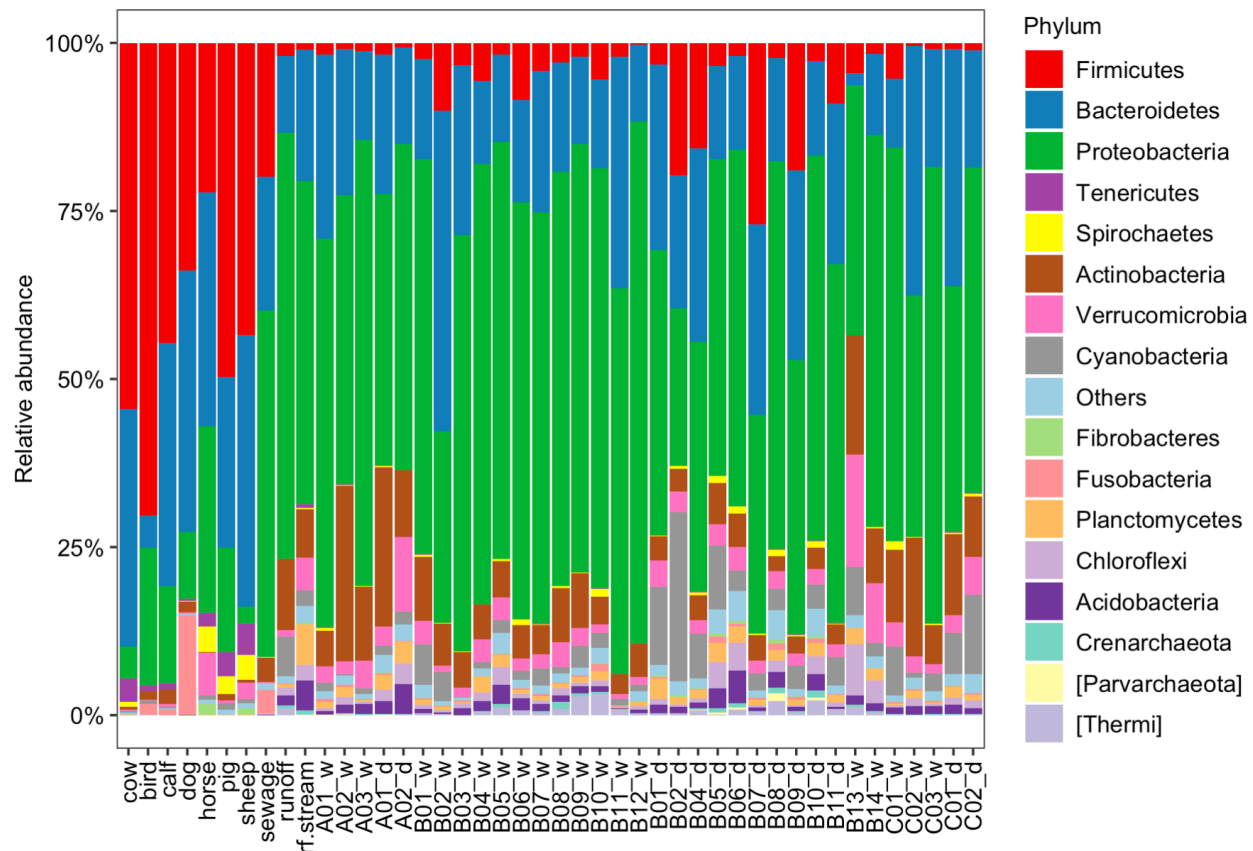


Figure 6-2 Microbial population of streams and possible pollution sources at phylum level.

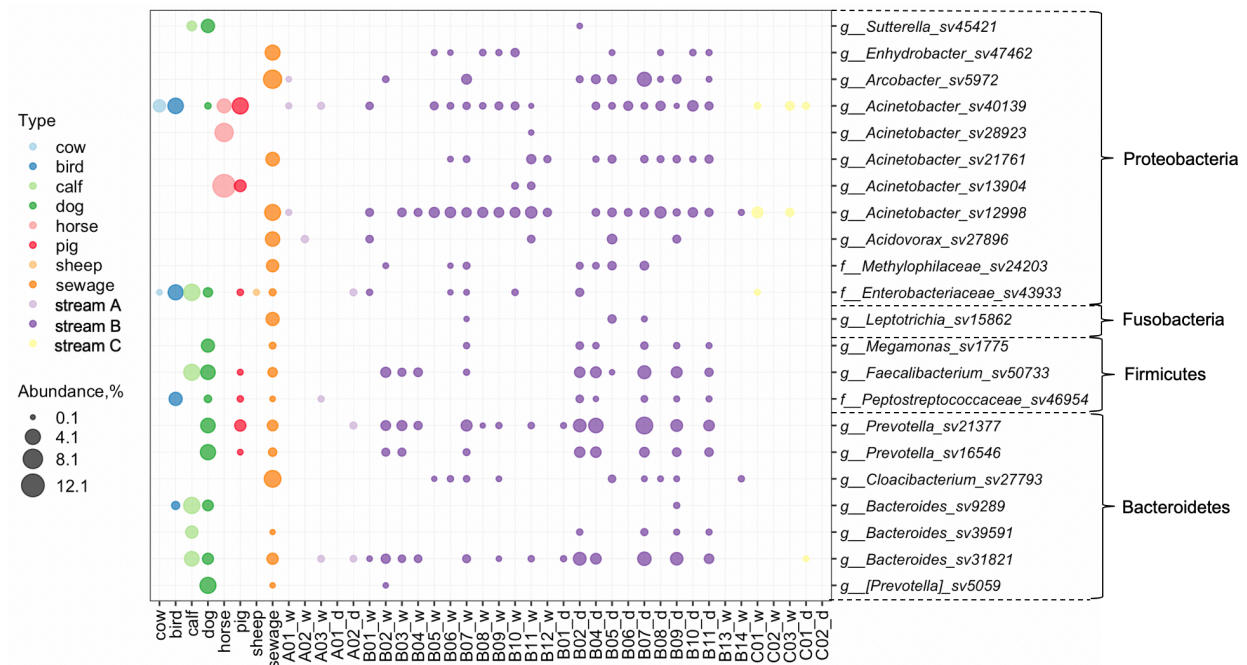


Figure 6-3 Bubble plots depicts the relative abundances of the main fecal sequences.

The values were the mean value of the replicates. Row represents individual sequence, and the sequences were grouped by phylum with brackets '}'. The lowest assigned taxonomy of each sequence was shown on the right side of the figure.

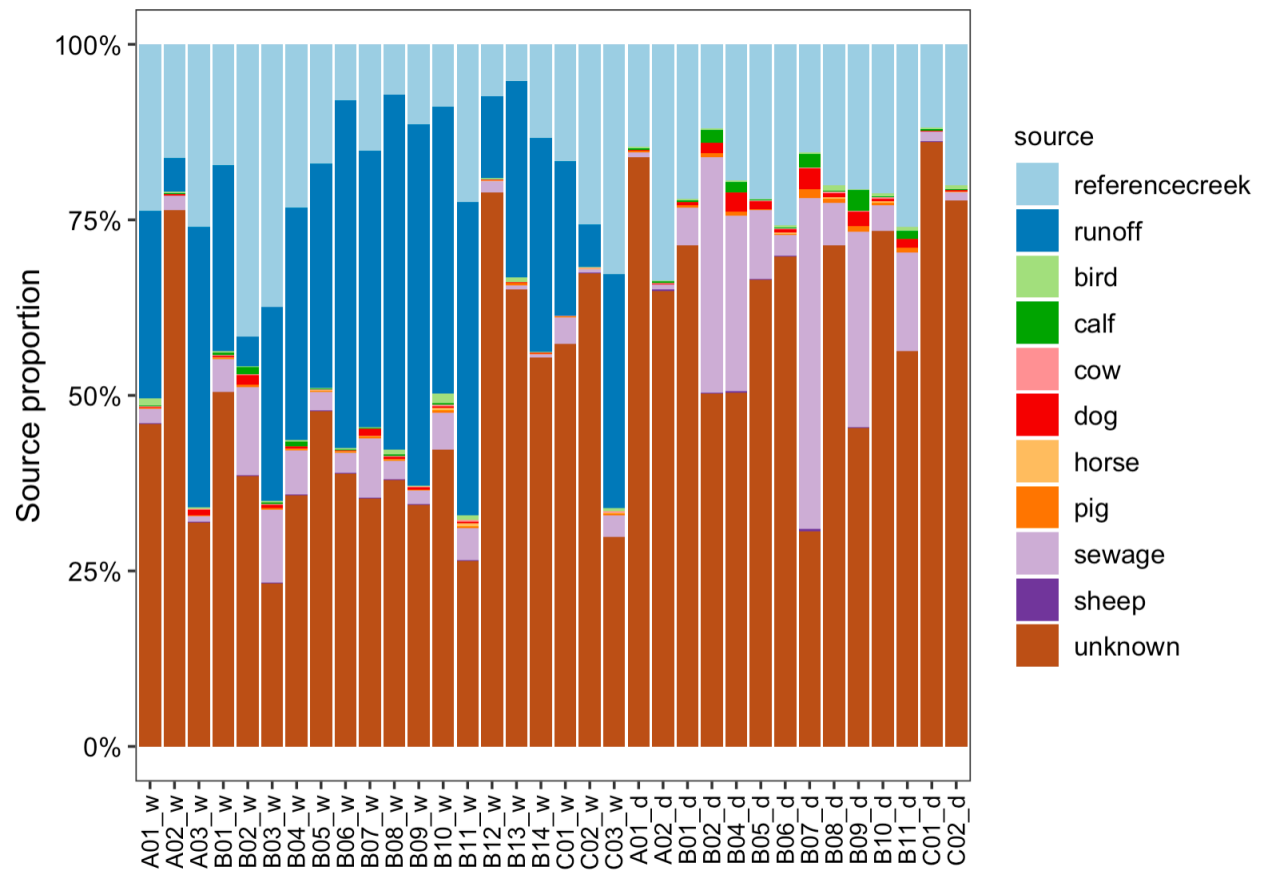


Figure 6-5 Source contribution to urban streams estimated by SourceTracker.

Animal feces are labeled as animal names.

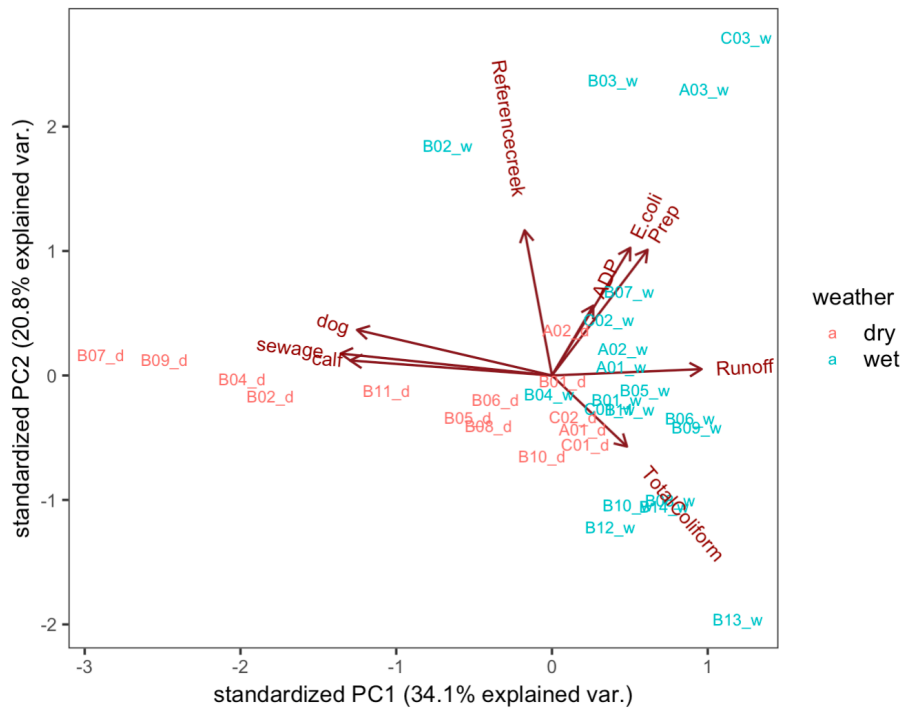


Figure 6-6 Principle component analysis of samples relatedness, environmental variables and source contributions.

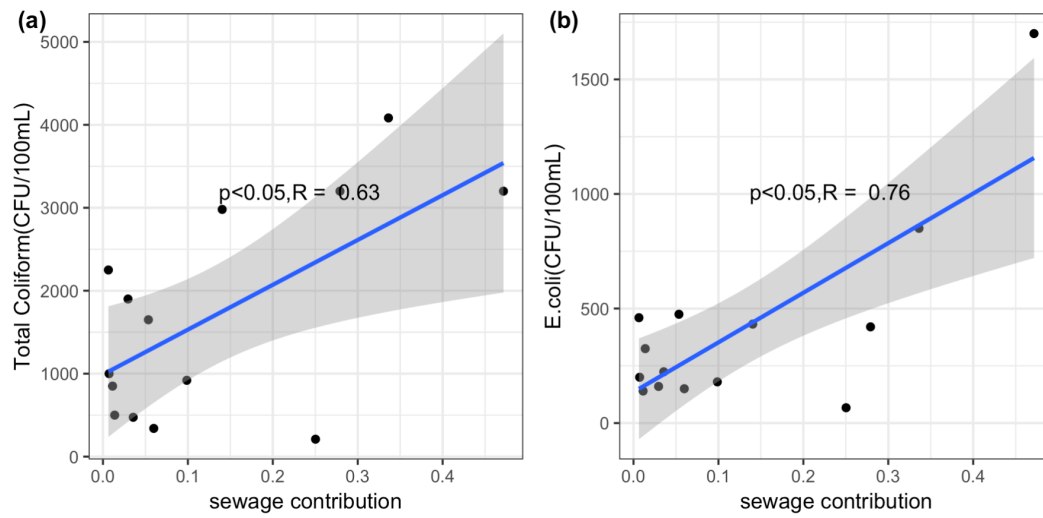


Figure 6-7 Correlation between sewage contribution and coliforms in baseline.

p stands for p-value of the significant test, R stands for correlation coefficients.

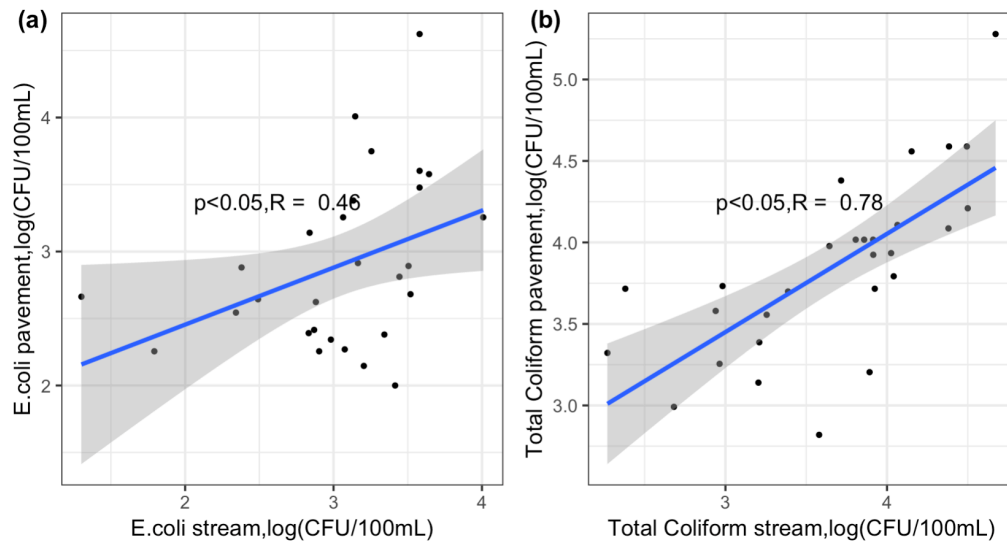


Figure 6-8 Correlation between sewage contribution and coliforms in baseline.

p stands for p-value of the significant test, R stands for correlation coefficients.

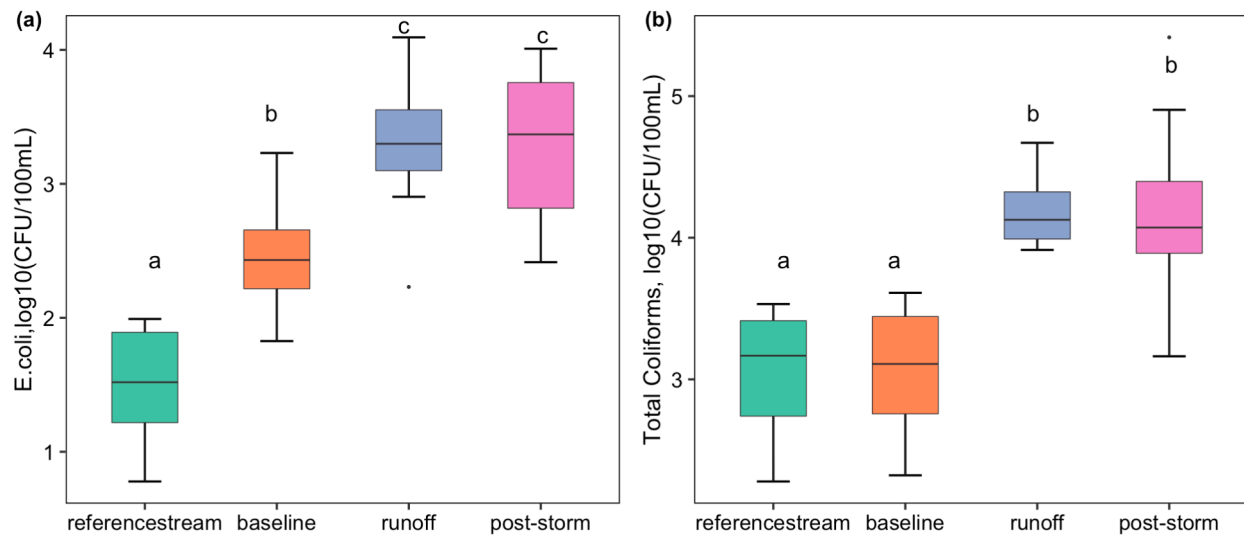


Figure 6-9 Coliform concentration in waterbody and stormwater runoff.

Boxplot with error bars indicating the standard deviations. Coliforms in different environmental water labeled with different letters indicate the significant differences with the other environmental water (ANOVA, Tukey HSD test, $p < 0.05$).

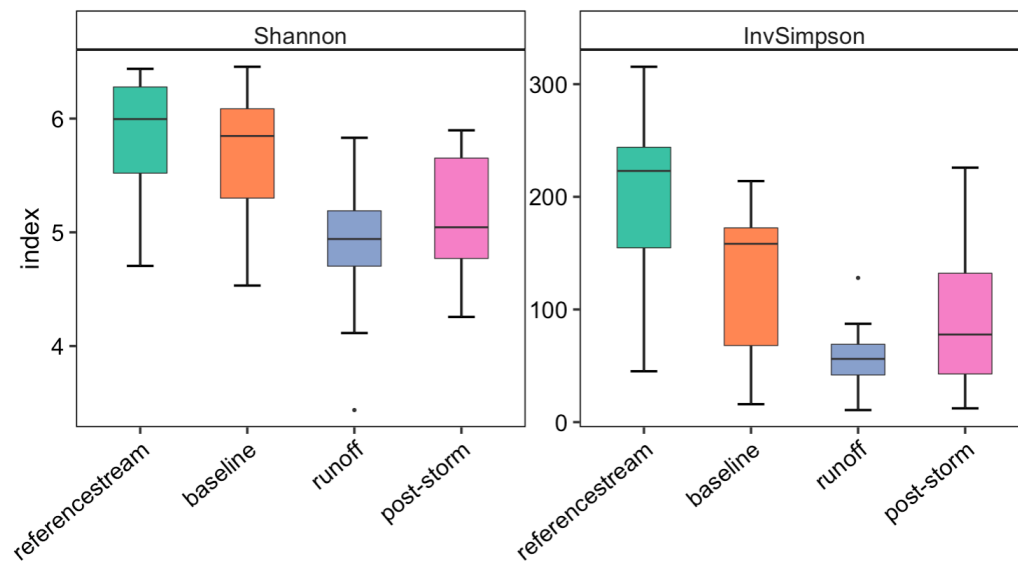


Figure 6-10 Alpha diversity of the microbial community in waterbody and stormwater runoff.

Boxplot with error bars indicating the standard deviations

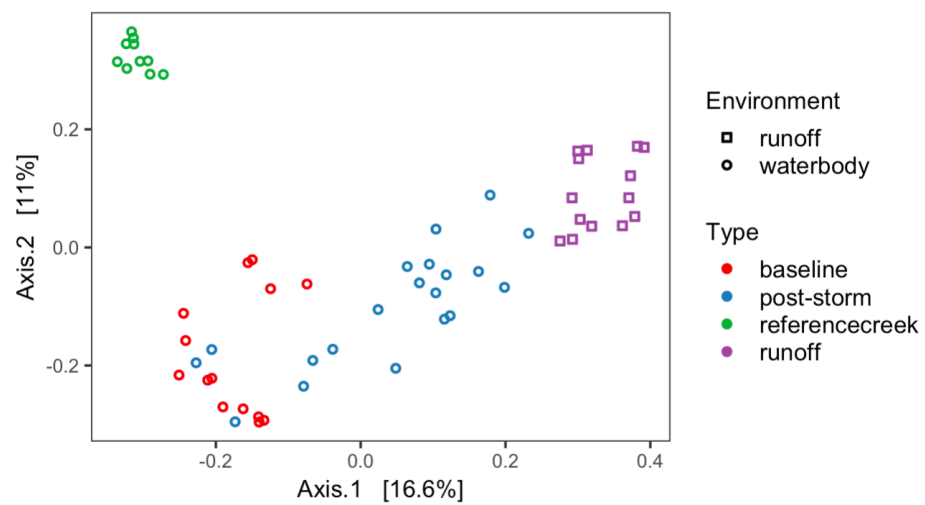


Figure 6-11 Sample relatedness of stormwater samples and stream water samples.

NMDS analysis with Unifrac dissimilarity matrix describes the relatedness among samples.

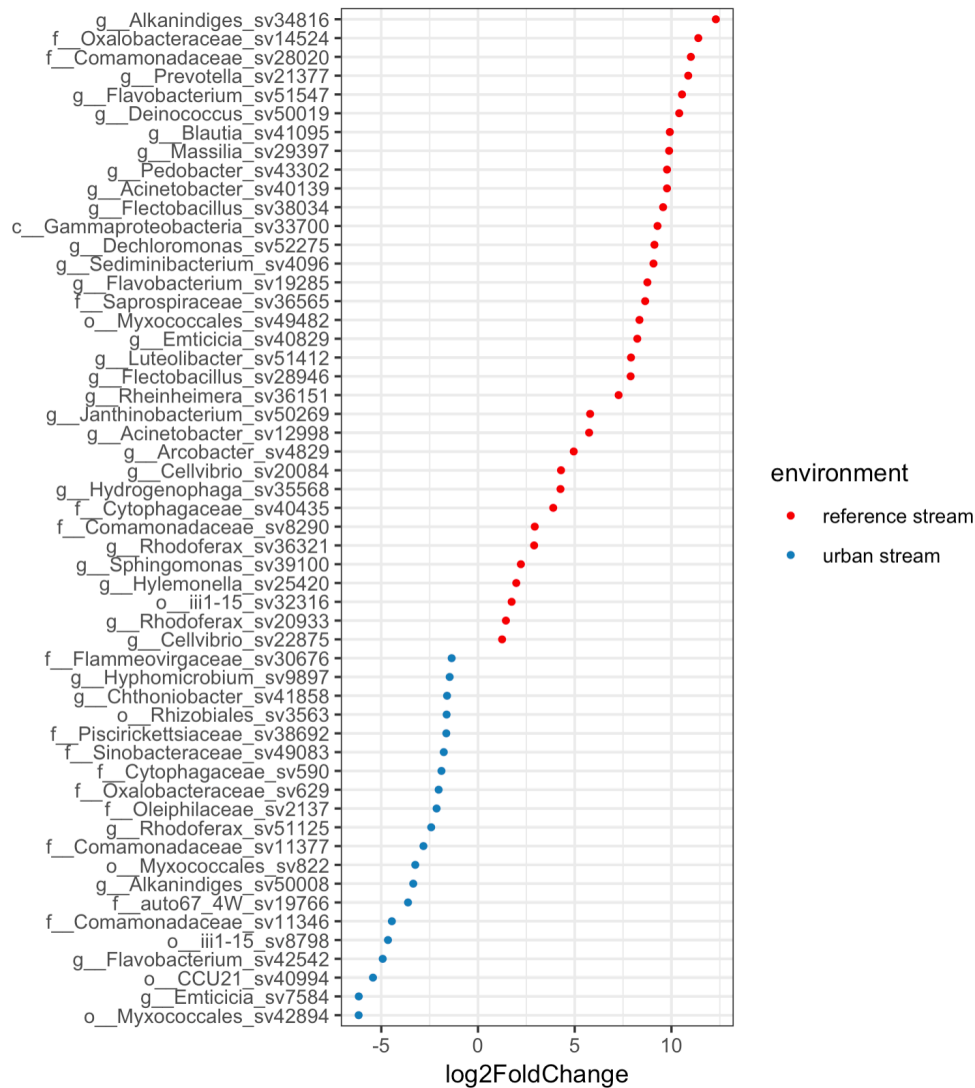


Figure 6-12 Log2Fold changes of taxonomic sequence variants between urban streams and reference streams.

The abundances of the sequence variants in the 16S rRNA gene amplicon data were used to calculate the log2FoldChange value by DESeq function in DESeq2 package in R. A positive log2 fold change for a comparison of urban stream with reference stream represents relative abundance in urban streams is larger in comparison to reference streams. Rare sequence variants were trimmed with sequence variants with relative abundance greater than 1% and existed in more than 10 samples, DESeq analysis result was filtered with p-value <0.01.

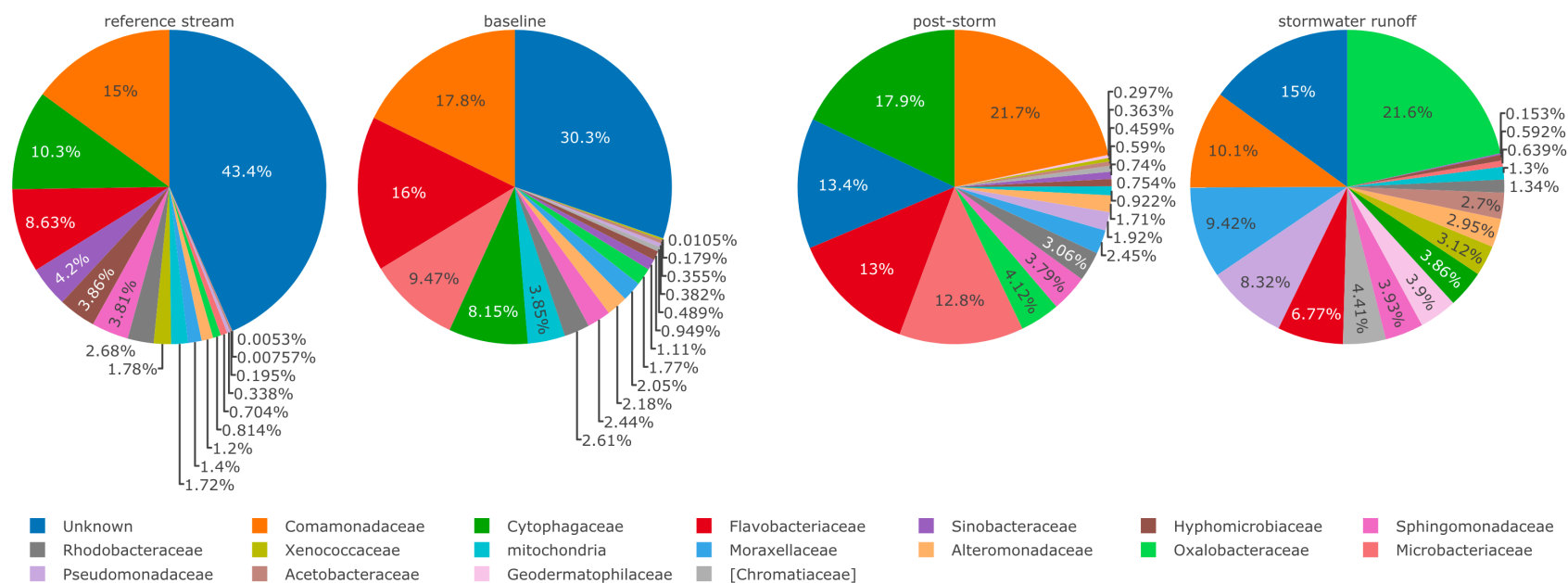


Figure 6-13 Microbial community of streams and stormwater runoff at the family level.

Families with relative abundance greater than 1% and exist in at least one type of sample were shown in the plot, while others were collapsed to 'other' category.

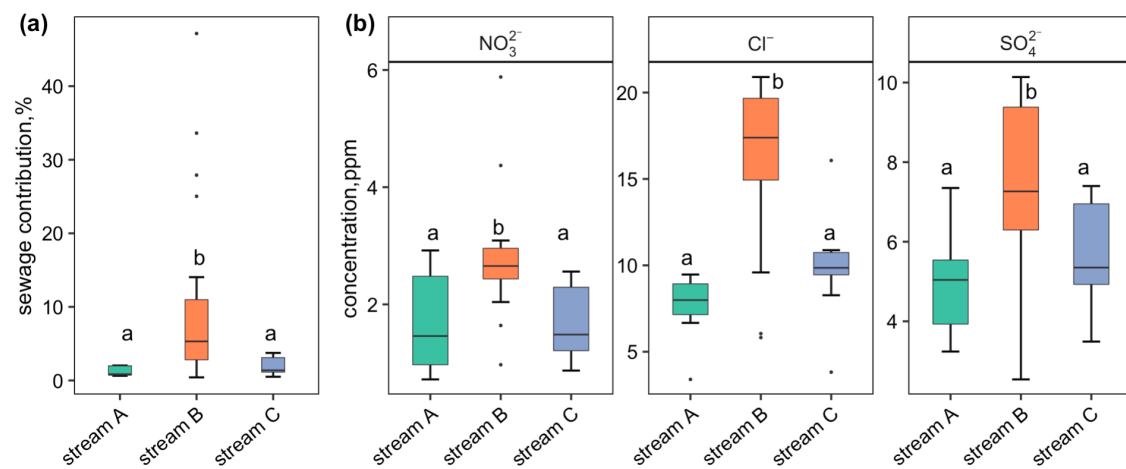


Figure 6-14 Site comparison of sewage inputs and nutrients concentration in urban streams.

Bars labeled with different labels indicate the significant differences with the other environmental water (ANOVA, Tukey HSD test, $p < 0.05$).

CHAPTER 5

CONCLUSIONS

This dissertation applied microbial community analysis to identify populations relevant to pollutant mitigation in natural and engineered processes. We discovered the distribution of persistence of antibiotic resistant genes and their potential carriers following anaerobic treatment of animal wastewater. Besides, microbial populations response to environmental disturbance, represented by the amendment of formate as additional substrate, were found. It explained the microbial assembly mechanism in animal wastewater treatment. Furthermore, populations highly tolerance to toxic wastewater could be developed in AnMBR. Microbial community analysis helps improve the treatment efficiency in recalcitrant industrial wastewater. Moreover, we discovered the dynamics of microbial community distribution of sewage. Finally, we find that microbial community analysis was capable of identifying sources of impaired surface water.

Chapter 2 successfully discovered that 126 ARG groups from more than 20 ARG categories persisted under engineered anaerobic treatment, mainly expressed in the most abundant unchanged subtypes, *vanUG*, *dfpE*, *msrB*, *PmrE*, *LlmA*, which belonged to resistant types, *macrolide-lincosamide-streptogramin*, *glycopeptide*, and *multidrug*, etc. The existence of host bacteria accompanied such persistence of ARGs, mainly originated from *Firmicutes*, *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes*. The persistence was supported by the persistent existence of some resistant functions and mobile genetic elements (MGEs).

Within Chapter 3, section A tested the hypothesis that formate would be favorable to a group of hydrogenotrophic methanogens under long term anaerobic treatment. Interestingly, the results rejected the hypothesis. In contrast to the hypothesized hydrogenotrophic methanogens, acetoclastic methanogens were surprisingly the dominant methanogens in balanced engineered anaerobic processes. Such phenomena could be attributed to homoacetogens, which likely helped *Methanosaeta* convert proposed hydrogenotrophic methanogenesis pathway to the acetoclastic pathway. Those possible functioned homoacetogens and acetoclastic methanogens were identified as ‘home’ populations that involve in legacy effects. The “home” bacteria and “home” methanogens coupled and led the pathway redirection under the formate treatment. This finding provides valuable insights into developing strategies for animal wastewater mitigation.

Within Chapter 3, section B extended the pollution mitigation question to industrial pollutants, recalcitrant biofuel process wastewater. The main finding was that a microbial consortium could recover methane from carbon-enriched (COD concentration exceeding 15,000 mg/L) wastewater. With the gradual increase in organic loading of the process wastewater, the AnMBR was able to remove >99% of COD loading with stable biogas production, which coincided with the dominance of *Methanobacterium*, suggesting the functional importance of hydrogenotrophic methanogenesis in the anaerobic metabolism of the wastewater organics. Other methanogenic populations, including *Methanosaeta* and *Methanomassiliicoccus*, were also found at significant levels in the AnMBR, indicative of the need for functional diversity to support the degradation of diverse organic constituents in the process wastewater. More importantly, the prevalence of multiple syntrophic bacteria (*Syntrophaceticus*, *Syntrophobacter*, *Syntrophomonas*, *Syntrophorhabdus*, and *Syntrophus*) demonstrates the importance of syntrophic interactions in anaerobic treatment. This finding innovatively designs and demonstrates a new path into the utilization of ‘recalcitrant’ industrial wastewater.

Within Chapter 4, section A experimental evidence implicated sewage from the pump station, not 24-hour composite sewage, is better microbiome assemblages in source tracking. Results showing that a subset of human fecal microbiome rapidly decayed in raw sewage, a group of non-fecal microbiome propagated in sewage conveyance system, and a group of human fecal microbiome thrived with sewage aging. It was also supported by the increased microbiome diversity with the aging and transportation of sewage in the natural environment. Non-overlaps of network hubs, nodes, and connectors were found among constructed RMT networks. This work renews the current understanding of sewage in source tracking studies to date.

Within Chapter 4, section B applied the results from part one into source tracking application. Natural stream health was assessed by microbial community composition. The health state was determined by the degree of human perturbations, which brought about the variation in diverse source contribution. In contrast, reference streams explicated the low variation in microbial community composition and high phylogenetic diversity, which preserved more pristine microbial profile in streams. This study verified an ‘Anna Karenina principle’ for stream

water microbiomes, in which urban stream water vary more in microbial community composition than pristine streams—similar to Leo Tolstoy's dictum that “all happy families look alike; each unhappy family is unhappy in its own way.”

In conclusion, this dissertation supports the assumption that the microbial population can help identify and evaluate the pollution sources, helps improve the pollution treatment process as well. These microbial populations may be novel indicators in pollutants identification and mitigation for waste treatment and management.

LIST OF REFERENCES

- Ahmed, W., Staley, C., Sadowsky, M.J., Gyawali, P., Sidhu, J.P.S., Palmer, A., Beale, D.J. and Toze, S. (2015) Toolbox Approaches Using Molecular Markers and 16S rRNA Gene Amplicon Data Sets for Identification of Fecal Pollution in Surface Water. *Applied and Environmental Microbiology* 81(20), 7067-7077.
- Alm, E.W., Burke, J. and Spain, A. (2003) Fecal indicator bacteria are abundant in wet sand at freshwater beaches. *Water research* 37(16), 3978-3982.
- Amaranthus, M. and Perry, D. (1989) Interaction effects of vegetation type and Pacific madrone soil inocula on survival, growth, and mycorrhiza formation of Douglas-fir. *Canadian Journal of Forest Research* 19(5), 550-556.
- Anderson, M.L., Whitlock, J.E. and Harwood, V.J. (2005) Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied and Environmental Microbiology* 71(6), 3041-3048.
- Angelidaki, I. and Sanders, W. (2004) Assessment of the anaerobic biodegradability of macropollutants. *Reviews in Environmental Science and Biotechnology* 3(2), 117-129.
- Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P. and Zhang, L. (2018a) DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 6(1), 23.
- Arango-Argoty, G., Garner, E., Prudent, A., Heath, L.S., Vikesland, P. and Zhang, L.Q. (2018b) DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 6.
- Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C. and Segata, N. (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3.
- Bae, S. and Wuertz, S. (2009) Rapid decay of host-specific fecal Bacteroidales cells in seawater as measured by quantitative PCR with propidium monoazide. *Water research* 43(19), 4850-4859.
- Balleste, E. and Blanch, A.R. (2010) Persistence of Bacteroides species populations in a river as measured by molecular and culture techniques. *Appl Environ Microbiol* 76(22), 7608-7616.
- Baral, D., Speicher, A., Dvorak, B., Admiraal, D. and Li, X. (2018) Quantifying the Relative Contributions of Environmental Sources to the Microbial Community in an Urban Stream under Dry and Wet Weather Conditions. *Applied and Environmental Microbiology* 84(15).
- Barbosa, T.M. and Levy, S.B. (2000) The impact of antibiotic use on resistance development and persistence. *Drug Resistance Updates* 3(5), 303-311.
- Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H. and Vavilin, V. (2002) The IWA anaerobic digestion model no 1 (ADM1). *Water Science and Technology* 45(10), 65-73.
- Bauerle, M.R., Schwalm, E.L. and Booker, S.J. (2015) Mechanistic Diversity of Radical S-Adenosylmethionine (SAM)-dependent Methylation. *Journal of Biological Chemistry* 290(7), 3995-4002.
- Becker, K., van Alen, S., Idelevich, E.A., Schleimer, N., Seggewiss, J., Mellmann, A., Kaspar, U. and Peters, G. (2018) Plasmid-Encoded Transferable mecB-Mediated Methicillin Resistance in *Staphylococcus aureus*. *Emerging Infectious Diseases* 24(2), 242-248.
- Bernhard, A.E. and Field, K.G. (2000) A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Applied and Environmental Microbiology* 66(10), 4571-4574.
- Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J. and Deneff, V.J. (2017) Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. *Environ Microbiol* 19(3), 1149-1162.
- Black, B.A., Michener, W.E., Ramirez, K.J., Bidby, M.J., Knott, B.C., Jarvis, M.W., Olstad, J., Mante, O.D., Dayton, D.C. and Beckham, G.T. (2016) Aqueous Stream Characterization from Biomass Fast Pyrolysis and Catalytic Fast Pyrolysis. *Acs Sustainable Chemistry & Engineering* 4(12), 6815-6827.

- Blake, L.I., Halim, F.A., Gray, C., Mair, R., Manning, D.A.C., Sallis, P., Hutchinson, H. and Gray, N.D. (2017) Evaluating an anaerobic digestion (AD) feedstock derived from a novel non-source segregated municipal solid waste (MSW) product. *Waste Management* 59, 149-159.
- Boechat, I.G., Kruger, A., Giani, A., Figueredo, C.C. and Gucker, B. (2011) Agricultural land-use affects the nutritional quality of stream microbial communities. *FEMS microbiology ecology* 77(3), 568-576.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114-2120.
- Boone, D.R. and Bryant, M.P. (1980) Propionate-Degrading Bacterium, *Syntrophobacter-Wolinii* Sp-Nov Gen-Nov from Methanogenic Ecosystems. *Applied and Environmental Microbiology* 40(3), 626-632.
- Boone, D.R., Johnson, R.L. and Liu, Y. (1989) Diffusion of the interspecies electron carriers H₂ and formate in methanogenic ecosystems and its implications in the measurement of K_m for H₂ or formate uptake. *Applied and Environmental Microbiology* 55(7), 1735-1741.
- Bovio, P., Cabezas, A. and Etchebehere, C. (2019) Preliminary analysis of Chloroflexi populations in full-scale UASB methanogenic reactors. *J Appl Microbiol* 126(2), 667-683.
- Bradshaw, J.K., Snyder, B.J., Oladeinde, A., Spidle, D., Berrang, M.E., Meinersmann, R.J., Oakley, B., Sidle, R.C., Sullivan, K. and Molina, M. (2016) Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. *Water research* 101, 498-509.
- Brauner, A., Fridman, O., Gefen, O. and Balaban, N.Q. (2016) Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nature Reviews Microbiology* 14(5), 320-330.
- Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L. and Sadowsky, M.J. (2017) A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary. *Environmental Science & Technology* 51(15), 8263-8271.
- Burton, G.A., Gunnison, D. and Lanza, G.R. (1987) Survival of Pathogenic Bacteria in Various Fresh-Water Sediments. *Applied and Environmental Microbiology* 53(4), 633-638.
- Campanaro, S., Treu, L., Kougias, P.G., De Francisci, D., Valle, G. and Angelidaki, I. (2016) Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. *Biotechnology for Biofuels* 9.
- Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., Gordon, J.I. and Knight, R. (2011) Moving pictures of the human microbiome. *Genome Biology* 12(5).
- Carrillo, M., Estrada, E. and Hazen, T.C. (1985) Survival and Enumeration of the Fecal Indicators *Bifidobacterium-Adolescentis* and *Escherichia-Coli* in a Tropical Rain-Forest Watershed. *Applied and Environmental Microbiology* 50(2), 468-476.
- Chen, B.W., Yang, Y., Liang, X.M., Yu, K., Zhang, T. and Li, X.D. (2013) Metagenomic Profiles of Antibiotic Resistance Genes (ARGs) between Human Impacted Estuary and Deep Ocean Sediments. *Environmental Science & Technology* 47(22), 12753-12760.
- Chen, Q.L., An, X.L., Li, H., Su, J.Q., Ma, Y.B. and Zhu, Y.G. (2016a) Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. *Environment International* 92-93, 1-10.
- Chen, S., Cheng, H., Liu, J., Hazen, T.C., Huang, V. and He, Q. (2017) Unexpected competitiveness of *Methanosaeta* populations at elevated acetate concentrations in methanogenic treatment of animal wastewater. *Appl Microbiol Biotechnol* 101(4), 1729-1738.
- Chen, S., Cheng, H., Wyckoff, K.N. and He, Q. (2016b) Linkages of Firmicutes and Bacteroidetes populations to methanogenic process performance. *J Ind Microbiol Biotechnol* 43(6), 771-781.

- Chen, S., Cheng, H.C., Wyckoff, K.N. and He, Q. (2016c) Linkages of Firmicutes and Bacteroidetes populations to methanogenic process performance. *Journal of Industrial Microbiology & Biotechnology* 43(6), 771-781.
- Chen, S., Zamudio Canas, E.M., Zhang, Y., Zhu, Z. and He, Q. (2012) Impact of substrate overloading on archaeal populations in anaerobic digestion of animal waste. *J Appl Microbiol* 113(6), 1371-1379.
- Chen, Y., Cheng, J.J. and Creamer, K.S. (2008) Inhibition of anaerobic digestion process: A review. *Bioresource technology* 99(10), 4044-4064.
- Cho, K.H., Pachepsky, Y.A., Kim, J.H., Guber, A.K., Shelton, D.R. and Rowland, R. (2010) Release of *Escherichia coli* from the bottom sediment in a first-order creek: Experiment and reach-specific modeling. *Journal of Hydrology* 391(3-4), 322-332.
- Cobbold, R.N., Rice, D.H., Davis, M.A., Besser, T.E. and Hancock, D.D. (2006) Long-term persistence of multi-drug-resistant *Salmonella enterica* serovar Newport in two dairy herds. *Javma-Journal of the American Veterinary Medical Association* 228(4), 585-591.
- Conklin, A., Stensel, H.D. and Ferguson, J. (2006) Growth kinetics and competition between *Methanosarcina* and *Methanosaeta* in mesophilic anaerobic digestion. *Water Environment Research* 78(5), 486-496.
- Conrad, R. and Wetter, B. (1990) Influence of temperature on energetics of hydrogen metabolism in homoacetogenic, methanogenic, and other anaerobic bacteria. *Archives of Microbiology* 155(1), 94-98.
- Crable, B.R., Sieber, J.R., Mao, X., Alvarez-Cohen, L., Gunsalus, R., Loo, R.R.O., Nguyen, H. and McInerney, M.J. (2016) Membrane Complexes of *Syntrophomonas wolfei* Involved in Syntrophic Butyrate Degradation and Hydrogen Formation. *Frontiers in Microbiology* 7.
- Cydzik-Kwiatkowska, A. and Zielińska, M. (2016) Bacterial communities in full-scale wastewater treatment systems. *World Journal of Microbiology and Biotechnology* 32(4), 66.
- Czekalski, N., Sigdel, R., Birtel, J., Matthews, B. and Burgmann, H. (2015) Does human activity impact the natural antibiotic resistance background? Abundance of antibiotic resistance genes in 21 Swiss lakes. *Environment International* 81, 45-55.
- Dareioti, M.A. and Kornaros, M. (2015) Anaerobic mesophilic co-digestion of ensiled sorghum, cheese whey and liquid cow manure in a two-stage CSTR system: effect of hydraulic retention time. *Bioresource technology* 175, 553-562.
- De Vrieze, J., Saunders, A.M., He, Y., Fang, J., Nielsen, P.H., Verstraete, W. and Boon, N. (2015) Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. *Water research* 75, 312-323.
- Deng, Y., Jiang, Y.H., Yang, Y.F., He, Z.L., Luo, F. and Zhou, J.Z. (2012) Molecular ecological network analyses. *BMC Bioinformatics* 13.
- Depardieu, F., Mejean, V. and Courvalin, P. (2015) Competition between VanU(G) Repressor and VanR(G) Activator Leads to Rheostatic Control of vanG Vancomycin Resistance Operon Expression. *PloS Genetics* 11(4).
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72(7), 5069-5072.
- Diekert, G. and Wohlfarth, G. (1994) Metabolism of homoacetogens. *Antonie van Leeuwenhoek* 66(1), 209-221.
- Dieppois, G., Ducret, V., Caille, O. and Perron, K. (2012) The Transcriptional Regulator CzcR Modulates Antibiotic Resistance and Quorum Sensing in *Pseudomonas aeruginosa*. *PloS one* 7(5).
- Dridi, B., Fardeau, M.L., Ollivier, B., Raoult, D. and Drancourt, M. (2012) *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int J Syst Evol Microbiol* 62(Pt 8), 1902-1907.

Dvory, N.Z., Kuznetsov, M., Livshitz, Y., Gasser, G., Pankratov, I., Lev, O., Adar, E. and Yakirevich, A. (2018) Modeling sewage leakage and transport in carbonate aquifer using carbamazepine as an indicator. *Water research* 128, 157-170.

Ebentier, D.L., Hanley, K.T., Cao, Y.P., Badgley, B.D., Boehm, A.B., Ervin, J.S., Goodwin, K.D., Gourmelon, M., Griffith, J.F., Holden, P.A., Kelty, C.A., Lozach, S., McGee, C., Peed, L.A., Raith, M., Ryu, H., Sadowsky, M.J., Scott, E.A., Domingo, J.S., Schriewer, A., Sinigalliano, C.D., Shanks, O.C., Van De Werfhorst, L.C., Wang, D., Wuertz, S. and Jay, J.A. (2013) Evaluation of the repeatability and reproducibility of a suite of qPCR-based microbial source tracking methods. *Water research* 47(18), 6839-6848.

Escalante-Semerena, J.C. and Wolfe, R.S. (1984) Formaldehyde oxidation and methanogenesis. *Journal of Bacteriology* 158(2), 721-726.

Esquivel-Elizondo, S., Parameswaran, P., Delgado, A.G., Maldonado, J., Rittmann, B.E. and Krajmalnik-Brown, R. (2016) Archaea and Bacteria Acclimate to High Total Ammonia in a Methanogenic Reactor Treating Swine Waste. *Archaea* 2016, 10.

Findlay, S. (2010) Stream microbial ecology. *Journal of the North American Benthological Society* 29(1), 170-181.

Fisher, J.C., Newton, R.J., Dila, D.K. and McLellan, S.L. (2015) Urban microbial ecology of a freshwater estuary of Lake Michigan. *Elementa-Science of the Anthropocene* 3.

Fraga, I., Charters, F.J., O'Sullivan, A.D. and Cochrane, T.A. (2016) A novel modelling framework to prioritize estimation of non-point source pollution parameters for quantifying pollutant origin and discharge in urban catchments. *Journal of Environmental Management* 167, 75-84.

Gajdacs, M., Spengler, G. and Urban, E. (2017) Identification and Antimicrobial Susceptibility Testing of Anaerobic Bacteria: Rubik's Cube of Clinical Microbiology? *Antibiotics-Basel* 6(4).

Gao, J., Liu, G.J., Li, H.P., Xu, L., Du, L.L. and Yang, B. (2016) Predictive functional profiling using marker gene sequences and community diversity analyses of microbes in full-scale anaerobic sludge digesters. *Bioprocess and Biosystems Engineering* 39(7), 1115-1127.

Gao, P., Gu, C.C., Wei, X., Li, X., Chen, H., Jia, H.Z., Liu, Z.H., Xue, G. and Ma, C.Y. (2017) The role of zero valent iron on the fate of tetracycline resistance genes and class 1 integrons during thermophilic anaerobic co-digestion of waste sludge and kitchen waste. *Water research* 111, 92-99.

Garbisu, C., Garaiyurrebaso, O., Lanzen, A., Alvarez-Rodriguez, I., Arana, L., Blanco, F., Smalla, K., Grohmann, E. and Alkorta, I. (2018) Mobile genetic elements and antibiotic resistance in mine soil amended with organic wastes. *Science of The Total Environment* 621, 725-733.

Garzio-Hadzick, A., Shelton, D.R., Hill, R.L., Pachepsky, Y.A., Guber, A.K. and Rowland, R. (2010) Survival of manure-borne *E. coli* in streambed sediment: Effects of temperature and sediment properties. *Water research* 44(9), 2753-2762.

Ginestet, C. (2011) ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society Series a-Statistics in Society* 174, 245-245.

Gnecco, I., Berretta, C., Lanza, L.G. and La Barbera, P. (2005) Storm water pollution in the urban environment of Genoa, Italy. *Atmospheric Research* 77(1-4), 60-73.

Gonzalez-Fernandez, C., Sialve, B. and Molinuevo-Salces, B. (2015) Anaerobic digestion of microalgal biomass: Challenges, opportunities and research needs. *Bioresour Technol* 198, 896-906.

Goodwin, S. and Zeikus, J.G. (1987) Ecophysiological adaptations of anaerobic bacteria to low pH: analysis of anaerobic digestion in acidic bog sediments. *Applied and Environmental Microbiology* 53(1), 57-64.

Gothwal, R. and Shashidhar, T. (2015) Antibiotic Pollution in the Environment: A Review. *Clean-Soil Air Water* 43(4), 479-489.

- Gucker, B., Boechat, I.G. and Giani, A. (2009) Impacts of agricultural land use on ecosystem structure and whole-stream metabolism of tropical Cerrado streams. *Freshwater Biology* 54(10), 2069-2085.
- Gujer, W. and Zehnder, A.J. (1983) Conversion processes in anaerobic digestion. *Water Science and Technology* 15(8-9), 127-167.
- Guo, J.H., Li, J., Chen, H., Bond, P.L. and Yuan, Z.G. (2017) Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water research* 123, 468-478.
- Hagglund, M., Backman, S., Macellaro, A., Lindgren, F., Borgmasts, E., Jacobsson, K., Dryselius, R., Stenberg, P., Sjödin, A., Forsman, M. and Ahlinder, J. (2018) Accounting for Bacterial Overlap Between Raw Water Communities and Contaminating Sources Improves the Accuracy of Signature-Based Microbial Source Tracking. *Frontiers in Microbiology* 9.
- Haugland, R.A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L. and Shanks, O.C. (2010) Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Systematic and Applied Microbiology* 33(6), 348-357.
- Hayes, J.R., Wagner, D.D., English, L.L., Carr, L.E. and Joseph, S.W. (2005) Distribution of streptogramin resistance determinants among *Enterococcus faecium* from a poultry production environment of the USA. *Journal of Antimicrobial Chemotherapy* 55(1), 123-126.
- He, L.Y., Ying, G.G., Liu, Y.S., Su, H.C., Chen, J., Liu, S.S. and Zhao, J.L. (2016) Discharge of swine wastes risks water quality and food safety: Antibiotics and antibiotic resistance genes, from swine sources to the receiving environments. *Environment International* 92-93, 210-219.
- He, Q. and Sanford, R.A. (2002) Induction characteristics of reductive dehalogenation in the ortho-halophenol-respiring bacterium, *Anaeromyxobacter dehalogenans*. *Biodegradation* 13(5), 307-316.
- HEISE, R., MÜLLER, V. and GOTTSCALK, G. (1992) Presence of a sodium-translocating ATPase in membrane vesicles of the homoacetogenic bacterium *Acetobacterium woodii*. *The FEBS Journal* 206(2), 553-557.
- Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'Brien, P., Deletic, A. and McCarthy, D. (2016) Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water research* 93, 242-253.
- Huang, X., Rippy, M.A., Mehring, A.S., Winfrey, B.K., Jiang, S.C. and Grant, S.B. (2018) Shifts in dissolved organic matter and microbial community composition are associated with enhanced removal of fecal pollutants in urban stormwater wetlands. *Water research* 137, 310-323.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C. and Bork, P. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* 44(D1), D286-D293.
- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11.
- Ibekwe, A.M., Ma, J.C. and Murinda, S.E. (2016) Bacterial community composition and structure in an Urban River impacted by different pollutant sources. *Science of The Total Environment* 566, 1176-1185.
- Jamieson, R.C., Joy, D.M., Lee, H., Kostaschuk, R. and Gordon, R.J. (2005) Resuspension of sediment-associated *Escherichia coli* in a natural stream. *Journal of Environmental Quality* 34(2), 581-589.
- Jeanneau, L., Solecki, O., Wery, N., Jarde, E., Gourmelon, M., Communal, P.Y., Jadas-Hecart, A., Caprais, M.P., Gruau, G. and Pourcher, A.M. (2012) Relative Decay of Fecal Indicator Bacteria and Human-Associated Markers: A Microcosm Study Simulating Wastewater Input into Seawater and Freshwater. *Environmental Science & Technology* 46(4), 2375-2382.

- Jia, B.F., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P.Y., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S.L., Wright, G.D. and McArthur, A.G. (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research* 45(D1), D566-D573.
- Jonsson, L.J., Alriksson, B. and Nilvebrant, N.O. (2013) Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnology for Biofuels* 6.
- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28(1), 27-30.
- Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B. and Wuertz, S. (2007) 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water research* 41(16), 3701-3715.
- Kim, K.R., Owens, G., Ok, Y.S., Park, W.K., Lee, D.B. and Kwon, S.I. (2012) Decline in extractable antibiotics in manure-based composts during composting. *Waste Management* 32(1), 110-116.
- Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R. and Kelley, S.T. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature methods* 8(9), 761-U107.
- Koskey, A.M., Fisher, J.C., Eren, A.M., Ponce-Terashima, R., Reis, M.G., Blanton, R.E. and McLellan, S.L. (2014) *Blautia* and *Prevotella* sequences distinguish human and animal fecal pollution in Brazil surface waters. *Environ Microbiol Rep* 6(6), 696-704.
- Kotsyurbenko, O. (2005) Trophic interactions in the methanogenic microbial community of low-temperature terrestrial ecosystems. *FEMS microbiology ecology* 53(1), 3-13.
- Kotsyurbenko, O.R., Glagolev, M.V., Nozhevnikova, A.N. and Conrad, R. (2001) Competition between homoacetogenic bacteria and methanogenic archaea for hydrogen at low temperature. *FEMS microbiology ecology* 38(2-3), 153-159.
- Kreader, C.A. (1998) Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Appl Environ Microbiol* 64(10), 4103-4105.
- Kumar, P. and Saroj, D.P. (2014) Water–energy–pollution nexus for growing cities. *Urban Climate* 10, 846-853.
- La Reau, A.J., Meier-Kolthoff, J.P. and Suen, G. (2016) Sequence-based analysis of the genus *Ruminococcus* resolves its phylogeny and reveals strong host association. *Microbial Genomics* 2(12).
- Laliberte, P. and Grimes, D.J. (1982) Survival of *Escherichia-Coli* in Lake Bottom Sediment. *Applied and Environmental Microbiology* 43(3), 623-628.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10(3).
- Latif, M.A., Mehta, C.M. and Batstone, D.J. (2015) Low pH anaerobic digestion of waste activated sludge for enhanced phosphorous release. *Water research* 81, 288-293.
- Lee, D.G., Roehrdanz, P.R., Feraud, M., Ervin, J., Anumol, T., Jia, A., Park, M., Tamez, C., Morelius, E.W., Gardea-Torresdey, J.L., Izbicki, J., Means, J.C., Snyder, S.A. and Holden, P.A. (2015) Wastewater compounds in urban shallow groundwater wells correspond to exfiltration probabilities of nearby sewers. *Water research* 85, 467-475.
- Lee, J., Kim, E., Han, G., Tongco, J.V., Shin, S.G. and Hwang, S. (2018) Microbial communities underpinning mesophilic anaerobic digesters treating food wastewater or sewage sludge: A full-scale study. *Bioresour Technol* 259, 388-397.

- Lekshmili, M., Ammini, P., Adjei, J., Sanford, L.M., Shrestha, U., Kumari, S. and Varela, M.F. (2018) Modulation of antimicrobial efflux pumps of the major facilitator superfamily in *Staphylococcus aureus*. *Aims Microbiology* 4(1), 1-18.
- Leven, L., Nyberg, K. and Schnurer, A. (2012) Conversion of phenols during anaerobic digestion of organic solid waste--a review of important microorganisms and impact of temperature. *Journal of Environmental Management* 95 Suppl, S99-103.
- Levican, A., Collado, L., Yustes, C., Aguilar, C. and Figueras, M.J. (2014) Higher water temperature and incubation under aerobic and microaerobic conditions increase the recovery and diversity of *Arcobacter* spp. from shellfish. *Appl Environ Microbiol* 80(1), 385-391.
- Levin, D.B., Pitt, L. and Love, M. (2004) Biohydrogen production: prospects and limitations to practical application. *International Journal of Hydrogen Energy* 29(2), 173-185.
- Levy, S.B. and Marshall, B. (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine* 10(12), S122-S129.
- Lew, B., Tarre, S., Beliaevski, M., Dosoretz, C. and Green, M. (2009) Anaerobic membrane bioreactor (AnMBR) for domestic wastewater treatment. *Desalination* 243(1-3), 251-257.
- Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K. and Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10), 1674-1676.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Proc, G.P.D. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), 2078-2079.
- Lim, Y.W., Lee, S.A., Kim, S.B., Yong, H.Y., Yeon, S.H., Park, Y.K., Jeong, D.W. and Park, J.S. (2005) Diversity of denitrifying bacteria isolated from Daejeon sewage treatment plant. *Journal of Microbiology* 43(5), 383-390.
- Lin, H.J., Peng, W., Zhang, M.J., Chen, J.R., Hong, H.C. and Zhang, Y. (2013) A review on anaerobic membrane bioreactors: Applications, membrane fouling and future perspectives. *Desalination* 314, 169-188.
- Lin, Q., De Vrieze, J., Li, C., Li, J., Li, J., Yao, M., Hedenec, P., Li, H., Li, T. and Rui, J. (2017) Temperature regulates deterministic processes and the succession of microbial interactions in anaerobic digestion process. *Water research* 123, 134-143.
- Litton, R.M., Ahn, J.H., Sercu, B., Holden, P.A., Sedlak, D.L. and Grant, S.B. (2010) Evaluation of Chemical, Molecular, and Traditional Markers of Fecal Contamination in an Effluent Dominated Urban Stream. *Environmental Science & Technology* 44(19), 7369-7375.
- Liu, B. and Pop, M. (2009) ARDB-Antibiotic Resistance Genes Database. *Nucleic Acids Research* 37, D443-D447.
- Liu, G., Zhang, Y., van der Mark, E., Magic-Knezev, A., Pinto, A., van den Bogert, B., Liu, W., van der Meer, W. and Medema, G. (2018) Assessing the origin of bacteria in tap water and distribution system in an unchlorinated drinking water system by SourceTracker using microbial community fingerprints. *Water research* 138, 86-96.
- Liu, R., Hao, X. and Wei, J. (2016) Function of homoacetogenesis on the heterotrophic methane production with exogenous H₂/CO₂ involved. *Chemical Engineering Journal* 284, 1196-1203.
- Liu, Y. and Whitman, W.B. (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Annals of the New York Academy of Sciences* 1125(1), 171-189.
- Liu, Y.C., Dong, Q. and Shi, H.C. (2015) Distribution and population structure characteristics of microorganisms in urban sewage system. *Applied microbiology and biotechnology* 99(18), 7723-7734.

- Lozano, C., Aspiroz, C., Lasarte, J.J., Gomez-Sanz, E., Zarazaga, M. and Torres, C. (2011) Dynamic of nasal colonization by methicillin-resistant *Staphylococcus aureus* ST398 and ST1 after mupirocin treatment in a family in close contact with pigs. *Comp Immunol Microbiol Infect Dis* 34(1), e1-7.
- Ma, L.P., Xia, Y., Li, B., Yang, Y., Li, L.G., Tiedje, J.M. and Zhang, T. (2016) Metagenomic Assembly Reveals Hosts of Antibiotic Resistance Genes and the Shared Resistome in Pig, Chicken, and Human Feces. *Environmental Science & Technology* 50(1), 420-427.
- Ma, Q., Qu, Y.Y., Shen, W.L., Zhang, Z.J., Wang, J.W., Liu, Z.Y., Li, D.X., Li, H.J. and Zhou, J.T. (2015) Bacterial community compositions of coking wastewater treatment plants in steel industry revealed by Illumina high-throughput sequencing. *Bioresource technology* 179, 436-443.
- Mao, C., Feng, Y., Wang, X. and Ren, G. (2015a) Review on research achievements of biogas from anaerobic digestion. *Renewable and Sustainable Energy Reviews* 45, 540-555.
- Mao, C., Zhang, T., Wang, X., Feng, Y., Ren, G. and Yang, G. (2017) Process performance and methane production optimizing of anaerobic co-digestion of swine manure and corn straw. *Scientific reports* 7(1), 9379.
- Mao, D.Q., Yu, S., Rysz, M., Luo, Y., Yang, F.X., Li, F.X., Hou, J., Mu, Q.H. and Alvarez, P.J.J. (2015b) Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water research* 85, 458-466.
- Maragkaki, A., Vasileiadis, I., Fountoulakis, M., Kyriakou, A., Lasaridi, K. and Manios, T. (2018) Improving biogas production from anaerobic co-digestion of sewage sludge with a thermal dried mixture of food waste, cheese whey and olive mill wastewater. *Waste Management* 71, 644-651.
- Martinez, J.L. (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science* 321(5887), 365-367.
- Mata-Alvarez, J., Mace, S. and Llabres, P. (2000) Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresource technology* 74(1), 3-16.
- Mattioli, M.C., Sassoubre, L.M., Russell, T.L. and Boehm, A.B. (2017) Decay of sewage-sourced microbial source tracking markers and fecal indicator bacteria in marine waters. *Water research* 108, 106-114.
- Mayer, R.E., Sofill-Mas, S., Egle, L., Reischer, G.H., Schade, M., Fernandez-Cassi, X., Fuchs, W., Mach, R.L., Lindner, G., Kirschner, A., Gaisbauer, M., Piringer, H., Blaschke, A.P., Girones, R., Zessnet, M., Sommer, R. and Farnleitner, A.H. (2016) Occurrence of human-associated *Bacteroidetes* genetic source tracking markers in raw and treated wastewater of municipal and domestic origin and comparison to standard and alternative indicators of faecal pollution. *Water research* 90, 265-276.
- McInerney, M.J., Bryant, M.P. and Pfennig, N. (1979) Anaerobic Bacterium That Degrades Fatty-Acids in Syntrophic Association with Methanogens. *Archives of Microbiology* 122(2), 129-135.
- McLellan, S.L. and Eren, A.M. (2014) Discovering new indicators of fecal pollution. *Trends in Microbiology* 22(12), 697-706.
- McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N. and Sogin, M.L. (2010a) Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology* 12(2), 378-392.
- McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N. and Sogin, M.L. (2010b) Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ Microbiol* 12(2), 378-392.
- McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N. and Sogin, M.L. (2010c) Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent (vol 12, pg 378, 2010). *Environmental Microbiology* 12(5), 1376-1376.
- McLellan, S.L. and Roguet, A. (2019) The unexpected habitat in sewer pipes for the propagation of microbial communities and their imprint on urban waters. *Curr Opin Biotechnol* 57, 34-41.

- Menzel, P., Ng, K.L. and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7.
- Molina, R. and Amaranthus, M. (1990) Rhizosphere biology: ecological linkages between soil processes, plant growth, and community dynamics.
- Molitor, B., Kirchner, K., Henrich, A.W., Schmitz, S. and Rosenbaum, M.A. (2016) Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Scientific reports* 6, 31518.
- Moore, P.R., Evenson, A., Luckey, T.D., McCoy, E., Elvehjem, C.A. and Hart, E.B. (1946) Use of Sulfasuxidine, Streptothricin, and Streptomycin in Nutritional Studies with the Chick. *Journal of Biological Chemistry* 165(2), 437-441.
- Mosbaek, F., Kjeldal, H., Mulat, D.G., Albertsen, M., Ward, A.J., Feilberg, A. and Nielsen, J.L. (2016) Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters by combined protein-based stable isotope probing and metagenomics. *Isme Journal* 10(10), 2405-2418.
- Myint, M., Nirmalakhandan, N. and Speece, R.E. (2007) Anaerobic fermentation of cattle manure: modeling of hydrolysis and acidogenesis. *Water Res* 41(2), 323-332.
- Nakatsu, C.H., Byappanahalli, M.N. and Nevers, M.B. (2019) Bacterial Community 16S rRNA Gene Sequencing Characterizes Riverine Microbial Impact on Lake Michigan. *Frontiers in Microbiology* 10.
- Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M. and Sogin, M.L. (2015) Sewage Reflects the Microbiomes of Human Populations. *MBio* 6(2).
- Newton, R.J., VandeWalle, J.L., Borchardt, M.A., Gorelick, M.H. and McLellan, S.L. (2011) Lachnospiraceae and Bacteroidales Alternative Fecal Indicators Reveal Chronic Human Sewage Contamination in an Urban Harbor. *Applied and Environmental Microbiology* 77(19), 6972-6981.
- Ng, K.K., Shi, X., Ong, S.L. and Ng, H.Y. (2016) Pyrosequencing reveals microbial community profile in anaerobic bio-entrapped membrane reactor for pharmaceutical wastewater treatment. *Bioresour Technol* 200, 1076-1079.
- Nie, Y., Liu, H., Du, G. and Chen, J. (2007) Enhancement of acetate production by a novel coupled syntrophic acetogenesis with homoacetogenesis process. *Process Biochemistry* 42(4), 599-605.
- Novaes, R.F. (1986) Microbiology of anaerobic digestion. *Water Science and Technology* 18(12), 1-14.
- Nowotny, J., Dodson, J., Fiechter, S., Gur, T.M., Kennedy, B., Macyk, W., Bak, T., Sigmund, W., Yamawaki, M. and Rahman, K.A. (2018) Towards global sustainability: Education on environmentally clean energy technologies. *Renewable & Sustainable Energy Reviews* 81, 2541-2551.
- Palatsi, J., Viñas, M., Guivernau, M., Fernandez, B. and Flotats, X. (2011) Anaerobic digestion of slaughterhouse waste: main process limitations and microbial community interactions. *Bioresource technology* 102(3), 2219-2227.
- Pan, X., Angelidaki, I., Alvarado-Morales, M., Liu, H., Liu, Y., Huang, X. and Zhu, G. (2016) Methane production from formate, acetate and H₂/CO₂; focusing on kinetics and microbial characterization. *Bioresource technology* 218, 796-806.
- Pennino, M.J., Kaushal, S.S., Mayer, P.M., Utz, R.M. and Cooper, C.A. (2016) Stream restoration and sewers impact sources and fluxes of water, carbon, and nutrients in urban watersheds. *Hydrology and Earth System Sciences* 20(8), 3419-3439.
- Perry, D., Molina, R. and Amaranthus, M. (1987) Mycorrhizae, mycorrhizospheres, and reforestation: current knowledge and research needs. *Canadian Journal of Forest Research* 17(8), 929-940.
- Petrucci, G., Gromaire, M.C., Shorshani, M.F. and Chebbo, G. (2014) Nonpoint source pollution of urban stormwater runoff: a methodology for source analysis. *Environmental Science and Pollution Research* 21(17), 10225-10242.

- Phelps, T. and Zeikus, J. (1984) Influence of pH on terminal carbon metabolism in anoxic sediments from a mildly acidic lake. *Applied and Environmental Microbiology* 48(6), 1088-1095.
- Rachbauer, L., Beyer, R., Bochmann, G. and Fuchs, W. (2017) Characteristics of adapted hydrogenotrophic community during biomethanation. *Science of The Total Environment* 595, 912-919.
- Racine, J.S. (2012) RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics* 27(1), 167-172.
- Renaux, A. and Consortium, U. (2018) UniProt: the universal protein knowledgebase (vol 45, pg D158, 2017). *Nucleic Acids Research* 46(5), 2699-2699.
- Rodriguez-Mozaz, S., Chamorro, S., Marti, E., Huerta, B., Gros, M., Sanchez-Melsio, A., Borrego, C.M., Barcelo, D. and Balcazar, J.L. (2015) Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water research* 69, 234-242.
- Romero-Güiza, M.S., Wahid, R., Hernández, V., Møller, H. and Fernández, B. (2017) Improvement of wheat straw anaerobic digestion through alkali pre-treatment: Carbohydrates bioavailability evaluation and economic feasibility. *Science of The Total Environment* 595, 651-659.
- Roslev, P. and Bukh, A.S. (2011) State of the art molecular markers for fecal pollution source tracking in water. *Applied microbiology and biotechnology* 89(5), 1341-1355.
- Rotaru, A.E., Shrestha, P.M., Liu, F.H., Shrestha, M., Shrestha, D., Embree, M., Zengler, K., Wardman, C., Nevin, K.P. and Lovley, D.R. (2014) A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to *Methanosaeta* for the reduction of carbon dioxide to methane. *Energy & Environmental Science* 7(1), 408-415.
- Ryan, P., Forbes, C. and Collieran, E. (2008) Investigation of the diversity of homoacetogenic bacteria in mesophilic and thermophilic anaerobic sludges using the formyltetrahydrofolate synthetase gene. *Water Sci Technol* 57(5), 675-680.
- Saady, N.M.C. (2013) Homoacetogenesis during hydrogen production by mixed cultures dark fermentation: unresolved challenge. *International Journal of Hydrogen Energy* 38(30), 13172-13191.
- Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and molecular biology reviews* 61(2), 262-280.
- Schink, B., Montag, D., Keller, A. and Müller, N. (2017) Hydrogen or formate: Alternative key players in methanogenic degradation. *Environmental Microbiology Reports* 9(3), 189-202.
- Schnürer, A., Schink, B. and Svensson, B.H. (1996) *Clostridium ultunense* sp. nov., a mesophilic bacterium oxidizing acetate in syntrophic association with a hydrogenotrophic methanogenic bacterium. *International Journal of Systematic and Evolutionary Microbiology* 46(4), 1145-1152.
- Sedano-Nunez, V.T., Boeren, S., Stams, A.J.M. and Plugger, C.M. (2018) Comparative proteome analysis of propionate degradation by *Syntrophobacter fumaroxidans* in pure culture and in coculture with methanogens. *Environmental Microbiology* 20(5), 1842-1856.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biology* 12(6).
- Shanks, O.C., Kelty, C.A., Sivaganesan, M., Varma, M. and Haugland, R.A. (2009) Quantitative PCR for Genetic Markers of Human Fecal Pollution. *Applied and Environmental Microbiology* 75(17), 5507-5513.
- Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L. and McLellan, S.L. (2013) Comparison of the Microbial Community Structures of Untreated Wastewaters from Different Geographic Locales. *Applied and Environmental Microbiology* 79(9), 2906-2913.
- Sharma, V.K., Johnson, N., Cizmas, L., McDonald, T.J. and Kim, H. (2016) A review of the influence of treatment strategies on antibiotic resistant bacteria and antibiotic resistance genes. *Chemosphere* 150, 702-714.

- Smith, K.S. and Ingram-Smith, C. (2007) Methanosaeta, the forgotten methanogen? *Trends Microbiol* 15(4), 150-155.
- Song, W., Wang, X.J., Gu, J., Zhang, S.Q., Yin, Y.N., Li, Y., Qian, X. and Sun, W. (2017) Effects of different swine manure to wheat straw ratios on antibiotic resistance genes and the microbial community structure during anaerobic digestion. *Bioresource technology* 231, 1-8.
- South, C., Hogsett, D. and Lynd, L. (1995) Modeling simultaneous saccharification and fermentation of lignocellulose to ethanol in batch and continuous reactors. *Enzyme and Microbial Technology* 17(9), 797-803.
- St-Pierre, B. and Wright, A.-D.G. (2013) Metagenomic analysis of methanogen populations in three full-scale mesophilic anaerobic manure digesters operated on dairy farms in Vermont, USA. *Bioresource technology* 138, 277-284.
- Stadler, L.B., Vela, J.D., Jain, S., Dick, G.J. and Love, N.G. (2018) Elucidating the impact of microbial community biodiversity on pharmaceutical biotransformation during wastewater treatment. *Microbial Biotechnology* 11(6), 995-1007.
- Staley, C., Kaiser, T., Gidley, M.L., Enochs, I.C., Jones, P.R., Goodwin, K.D., Sinigalliano, C.D., Sadowsky, M.J. and Chun, C.L. (2017) Differential Impacts of Land-Based Sources of Pollution on the Microbiota of Southeast Florida Coral Reefs. *Applied and Environmental Microbiology* 83(10).
- Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V.J., Brown, C.M. and Sadowsky, M.J. (2018) Application of SourceTracker for Accurate Identification of Fecal Pollution in Recreational Freshwater: A Double-Blinded Study. *Environmental Science & Technology* 52(7), 4207-4217.
- Stams, A.J. and Plugge, C.M. (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology* 7(8), 568.
- Stoeckel, D.M., Stelzer, E.A. and Dick, L.K. (2009) Evaluation of two spike-and-recovery controls for assessment of extraction efficiency in microbial source tracking studies. *Water research* 43(19), 4820-4827.
- Stojkovic, V., Noda-Garcia, L., Tawfik, D. and Fujimori, D. (2017) Antibiotic resistance evolved via inactivation of a ribosomal RNA methylating enzyme. *Abstracts of Papers of the American Chemical Society* 253.
- Sun, W., Gu, J., Wang, X.J., Qian, X. and Tuo, X.X. (2018) Impacts of biochar on the environmental risk of antibiotic resistance genes and mobile genetic elements during anaerobic digestion of cattle farm wastewater. *Bioresource technology* 256, 342-349.
- Sun, W.M., Xiao, E.Z., Krumins, V., Dong, Y.R., Xiao, T.F., Ning, Z.P., Chen, H.Y. and Xiao, Q.X. (2016) Characterization of the microbial community composition and the distribution of Fe-metabolizing bacteria in a creek contaminated by acid mine drainage. *Applied microbiology and biotechnology* 100(19), 8523-8535.
- Svensson, K., Kj r laug, O., Higgins, M.J., Linjordet, R. and Horn, S.J. (2018) Post-anaerobic digestion thermal hydrolysis of sewage sludge and food waste: Effect on methane yields, dewaterability and solids reduction. *Water research*.
- Svojitka, J., Dvorak, L., Studer, M., Straub, J.O., Fromelt, H. and Wintgens, T. (2017) Performance of an anaerobic membrane bioreactor for pharmaceutical wastewater treatment. *Bioresource technology* 229, 180-189.
- TDEC (2016) 303(d) LIST. CONSERVATION, T.D.O.E.A. (ed).
- Team, R.D.C. (2016) R: A language and environment for statistical computing, R Foundation for Sattistical Computing, Vienna, Austria.
- Thevenon, F., Adatte, T., Wildi, W. and Pote, J. (2012) Antibiotic resistant bacteria/genes dissemination in lacustrine sediments highly increased following cultural eutrophication of Lake Geneva (Switzerland). *Chemosphere* 86(5), 468-476.
- Thomsen, M.H., Thygesen, A. and Thomsen, A.B. (2009) Identification and characterization of fermentation inhibitors formed during hydrothermal treatment and following SSF of wheat straw. *Applied microbiology and biotechnology* 83(3), 447-455.

- Tien, Y.C., Li, B., Zhang, T., Scott, A., Murray, R., Sabourin, L., Marti, R. and Topp, E. (2017) Impact of dairy manure pre-application treatment on manure composition, soil dynamics of antibiotic resistance genes, and abundance of antibiotic-resistance genes on vegetables at harvest. *Science of The Total Environment* 581, 32-39.
- Van Bambeke, F., Van Laethem, Y., Courvalin, P. and Tulkens, P.M. (2004) Glycopeptide antibiotics from conventional molecules to new derivatives. *Drugs* 64(9), 913-936.
- Vass, M. and Langenheder, S. (2017) The legacy of the past: effects of historical processes on microbial metacommunities. *Aquatic Microbial Ecology* 79(1), 13-19.
- Vispute, T.P. and Huber, G.W. (2009) Production of hydrogen, alkanes and polyols by aqueous phase processing of wood-derived pyrolysis oils. *Green Chemistry* 11(9), 1433-1445.
- Wang, K., Yin, J., Shen, D. and Li, N. (2014) Anaerobic digestion of food waste for volatile fatty acids (VFAs) production with different types of inoculum: effect of pH. *Bioresource technology* 161, 395-401.
- Wang, P., Chen, B., Yuan, R.Q., Li, C.Q. and Li, Y. (2016) Characteristics of aquatic bacterial community and the influencing factors in an urban river. *Science of The Total Environment* 569, 382-389.
- Westerholm, M., Isaksson, S., Sun, L. and Schnürer, A. (2017) Microbial community ability to adapt to altered temperature conditions influences operating stability in anaerobic digestion. *Energy Procedia* 105, 895-900.
- Westerholm, M., Roos, S. and Schnurer, A. (2010) *Syntrophaceticus schinkii* gen. nov., sp nov., an anaerobic, syntrophic acetate-oxidizing bacterium isolated from a mesophilic anaerobic filter. *Fems Microbiology Letters* 309(1), 100-104.
- Wijesinghe, D.T.N., Dassanayake, K.B., Sommer, S.G., Scales, P. and Chen, D. (2017) 2 Biogas improvement by adding Australian zeolite during the anaerobic digestion of C: N ratio adjusted swine manure. Enhancement of methane generation by reducing nitrogen concentration during anaerobic digestion of swine manure, 103.
- Wirth, B., Reza, T. and Mumme, J. (2015) Influence of digestion temperature and organic loading rate on the continuous anaerobic treatment of process liquor from hydrothermal carbonization of sewage sludge. *Bioresource technology* 198, 215-222.
- Wolin, E., Wolin, M.J. and Wolfe, R. (1963) Formation of methane by bacterial extracts. *Journal of Biological Chemistry* 238(8), 2882-2886.
- Wu, L., Yang, Y., Chen, S., Zhao, M., Zhu, Z., Yang, S., Qu, Y., Ma, Q., He, Z. and Zhou, J. (2016a) Long-term successional dynamics of microbial association networks in anaerobic digestion processes. *Water research* 104, 1-10.
- Wu, L.W., Yang, Y.F., Chen, S., Zhao, M.X., Zhu, Z.W., Yang, S.H., Qu, Y.Y., Ma, Q., He, Z.L., Zhou, J.Z. and He, Q. (2016b) Long-term successional dynamics of microbial association networks in anaerobic digestion processes. *Water research* 104, 1-10.
- Xiao, B., Qin, Y., Zhang, W., Wu, J., Qiang, H., Liu, J. and Li, Y.-Y. (2018) Temperature-phased anaerobic digestion of food waste: A comparison with single-stage digestions based on performance and energy balance. *Bioresource technology* 249, 826-834.
- Xiao, K.-Q., Beulig, F., Kjeldsen, K.U., Jørgensen, B.B. and Risgaard-Petersen, N. (2017) Concurrent Methane Production and Oxidation in Surface Sediment from Aarhus Bay, Denmark. *Frontiers in Microbiology* 8(1198).
- Xu, J., Xu, Y., Wang, H.M., Guo, C.S., Qiu, H.Y., He, Y., Zhang, Y., Li, X.C. and Meng, W. (2015) Occurrence of antibiotics and antibiotic resistance genes in a sewage treatment plant and its effluent-receiving river. *Chemosphere* 119, 1379-1385.
- Yakirevich, A., Pachepsky, Y.A., Guber, A.K., Gish, T.J., Shelton, D.R. and Cho, K.H. (2013) Modeling transport of *Escherichia coli* in a creek during and after artificial high-flow events: Three-year study and analysis. *Water research* 47(8), 2676-2688.

- Yang, Z.G., d'Auriac, M.A., Goggins, S., Kasprzyk-Hordern, B., Thomas, K.V., Frost, C.G. and Estrela, P. (2015) A Novel DNA Biosensor Using a Ferrocenyl Intercalator Applied to the Potential Detection of Human Population Biomarkers in Wastewater. *Environmental Science & Technology* 49(9), 5609-5617.
- Yeh, K.H., Shun, C.T., Chen, C.L., Lin, J.T., Lee, W.J., Lee, P.H., Chen, Y.C. and Cheng, A.L. (1998) High expression of thymidylate synthase is associated with the drug resistance of gastric carcinoma to high dose 5-fluorouracil-based systemic chemotherapy. *Cancer* 82(9), 1626-1631.
- Yenigün, O. and Demirel, B. (2013) Ammonia inhibition in anaerobic digestion: a review. *Process Biochemistry* 48(5-6), 901-911.
- Yin, X.L., Jiang, X.T., Chai, B.L., Li, L.G., Yang, Y., Cole, J.R., Tiedje, J.M. and Zhang, T. (2018) ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* 34(13), 2263-2270.
- Zaneveld, J.R., McMinds, R. and Thurber, R.V. (2017) Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nature microbiology* 2(9).
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M. and Larsen, M.V. (2012) Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67(11), 2640-2644.
- Zhang, B., Xu, X. and Zhu, L. (2017a) Structure and function of the microbial consortia of activated sludge in typical municipal wastewater treatment plants in winter. *Sci Rep* 7(1), 17930.
- Zhang, C., Su, H., Baeyens, J. and Tan, T. (2014) Reviewing the anaerobic digestion of food waste for biogas production. *Renewable and Sustainable Energy Reviews* 38, 383-392.
- Zhang, J., Loh, K.-C., Lee, J., Wang, C.-H., Dai, Y. and Tong, Y.W. (2017b) Three-stage anaerobic co-digestion of food waste and horse manure. *Scientific reports* 7(1), 1269.
- Zhang, J.Y., Chen, M.X., Sui, Q.W., Wang, R., Tong, J. and Wei, Y.S. (2016) Fate of antibiotic resistance genes and its drivers during anaerobic co-digestion of food waste and sewage sludge based on microwave pretreatment. *Bioresource technology* 217, 28-36.
- Zhang, Q., He, X. and Yan, T. (2015a) Differential Decay of Wastewater Bacteria and Change of Microbial Communities in Beach Sand and Seawater Microcosms. *Environmental Science & Technology* 49(14), 8531-8540.
- Zhang, R., Anderson, E., Addy, M., Deng, X., Kabir, F., Lu, Q., Ma, Y., Cheng, Y., Liu, Y. and Chen, P. (2017c) An innovative intermittent-vacuum assisted thermophilic anaerobic digestion process for effective animal manure utilization and treatment. *Bioresource technology* 244, 1073-1080.
- Zhang, T., Yang, Y. and Pruden, A. (2015b) Effect of temperature on removal of antibiotic resistance genes by anaerobic digestion of activated sludge revealed by metagenomic approach. *Applied microbiology and biotechnology* 99(18), 7771-7779.
- Zhang, T., Zhang, X.X. and Ye, L. (2011) Plasmid Metagenome Reveals High Levels of Antibiotic Resistance Genes and Mobile Genetic Elements in Activated Sludge. *PloS one* 6(10).
- Zhou, J.Z., Bruns, M.A. and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology* 62(2), 316-322.
- Zhu, L.Q., Wu, S.L., Zhang, G.P. and Zhu, G.Q. (2008) The cellular receptors for infectious bursal disease virus. *African Journal of Biotechnology* 7(25), 4832-4835.
- Zhu, Y.G., Johnson, T.A., Su, J.Q., Qiao, M., Guo, G.X., Stedtfeld, R.D., Hashsham, S.A. and Tiedje, J.M. (2013) Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proceedings of the National Academy of Sciences of the United States of America* 110(9), 3435-3440.
- Zhu, Z., Hsueh, M.K. and He, Q. (2011) Enhancing biomethanation of municipal waste sludge with grease trap waste as a co-substrate. *Renewable Energy* 36(6), 1802-1807.

Ziganshin, A.M., Schmidt, T., Lv, Z., Liebetrau, J., Richnow, H.H., Kleinsteuber, S. and Nikolausz, M. (2016) Reduction of the hydraulic retention time at constant high organic loading rate to reach the microbial limits of anaerobic digestion in various reactor systems. *Bioresource technology* 217, 62-71.

VITA

Lu Yang entered the University of Tennessee, Knoxville via CSC DEEP 100-Ph.D. Program in August 2015, under the guidance of Dr. Qiang He at the Department of Civil and Environmental Engineering. She is expected to attain her Ph.D. degree in Environmental Engineering with a focus on Environmental Microbiology, together with M.S. degree in Statistics in August 2019.