

**METAGENOMIC INSIGHTS INTO MICROBIAL DIVERSITY AND RESISTANCE TO
ANTIBIOTICS IN WASTEWATER TREATMENT PLANTS**

By

Mariya Munir

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

Environmental Engineering - Doctor of Philosophy

2014

ABSTRACT

METAGENOMIC INSIGHTS INTO MICROBIAL DIVERSITY AND RESISTANCE TO ANTIBIOTICS IN WASTEWATER TREATMENT PLANTS

By

Mariya Munir

Our water environment is greatly impacted by the presence of microbial contaminants which is a great concern in terms of public health exposure. Full-scale conventional and state-of-the-art wastewater utilities have been found to release pathogens and antibiotic resistant bacteria in the environment. Management and minimization of microbial pathogens and resistant bacteria in wastewater treatment plants is critical since the spread of pathogens and antibiotic resistant genes in the environment poses a significant challenge to diverse aspects of our global community. The overall aim of this study is to provide metagenomic insights into bacterial, viral and phage diversity and resistance to antibiotics and metal compounds in wastewater utilities. Samples were collected from two different wastewater treatment systems, a conventional activated sludge utility and a membrane bioreactor (MBR), in Michigan. Metagenomic analyses were conducted on Illumina Miseq and Hiseq generated sequences using MGRAST and METAVIR analysis software. The findings suggest that there is a substantial shift in the phage community over the course of the activated sludge process. Phage populations are dynamic and phage DNA was associated with antibiotic resistant genes in wastewater. It was observed that there are differences in the abundance of functional genes related to resistance (antibiotic resistance and metal resistance) in different samples. Genes coding for antibiotic resistance were identified in all bacterial samples along with genes coding for resistance to metals. The MBR utility samples showed slightly higher number of hits for all the functional categories compared to conventional wastewater treatment samples. Diverse viral and bacterial human pathogens were

observed in treated wastewater samples. Diversity analysis does not provide quantitative data on pathogen loads or infectivity but it provides a list of potentially pathogenic viruses and bacteria that need to be considered during treatment management decisions. This study provided a bioinformatics approach for identifying microbial diversity in different wastewater treatment stages and technologies. The results of this work provide significant information that will contribute to sustainable wastewater management decisions.

I would like to dedicate my thesis to my beloved father Late Mr. Muniruddin Hasan, my mother Mrs Shabnam Mirza and my husband Mr. Azhar Ahamad, who inspired me, had faith in me and made me an achiever. I always admire them for their hardwork, intelligence and kindness. I am blessed to have them in my life.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my advisor and mentor Dr. Irene Xagorarakis for her constant guidance, encouragements, research ideas and support throughout my Ph.D program and research. A special thank to Dr. Terence Marsh for guiding me and helping me during my PhD research.

I would also like to thank all my committee members, Dr. Syed Hashsham and Dr. Alison Cupples for guidance and thoughts on my research. I also thank my labmates for their help and cooperation.

Words can not describe the support from my parents, my in-laws and family members. They have always supported me in my personal and professional career and have been there in every thick and thin moment. I thank them all for their endless Love. I also want to thank my friends for their continuous help throughout my stay at Michigan State University.

And special thanks to my husband Mr. Azhar Ahamad, my strength, my support, my courage. He has been extremely patient, caring and encouraging; making me realize this dream is possible and achievable. This has been truly a memorable journey and would not have been complete with out the support of everyone.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1	1
INTRODUCTION.....	1
REFERENCES.....	6
CHAPTER 2	10
PHAGE AND ANTIBIOTIC RESISTANCE GENES IN A CONVENTIONAL WASTEWATER TREATMENT PLANT	10
Abstract	10
1. Introduction.....	11
2. Materials and Methods	13
2.1. Sample Collection:	13
2.2. Sample Processing:.....	14
2.3. DNA Extraction:.....	15
2.4. Quantification:	15
2.5. Metagenomic analyses:.....	15
2.5.1. Blast analyses:.....	16
2.5.2. MG-RAST analyses:.....	16
3. Results and Discussion:.....	16
3.1. Metagenome sequencing and assembly results:	16
3.2. Phage Diversity:	18
3.3. Antibiotic Resistant Gene Diversity:.....	19
3.4. Concentration of ARGs:	20
4. Conclusions:.....	21
5. Acknowledgement:.....	22
APPENDIX	23
REFERENCES.....	32
CHAPTER 3	37
METAGENOMIC INSIGHTS INTO MICROBIAL RESISTANCE TO ANTIBIOTIC AND METAL COMPOUNDS IN WASTEWATER UTILITIES.....	37
Abstract	37
1. Introduction.....	38
2. Materials and Methods	41
2.1. Sample Collection:	41
2.2. Sample Processing:.....	41
2.3. Nucleic acid Extraction:	41
2.5. Metagenomic sequencing and analyses:.....	42
2.6. MG-RAST analyses:.....	42
2.7. Statistical analysis:	43
3. Results and Discussion:.....	44

3.1. Metagenome sequencing and assembly results:	44
3.2. Resistance to antibiotics: Antibiotic Resistant Gene Diversity:	46
3.3. Resistance to metal compounds:.....	48
4. Conclusions:	50
5. Acknowledgement:.....	51
APPENDIX	52
REFERENCES.....	62
CHAPTER 4	68
SCREENING FOR POTENTIAL VIRAL AND BACTERIAL PATHOGENS IN WASTEWATER EFFLUENT RELEASED FROM AN MBR AND A CONVENTIONAL TREATMENT UTILITY USING METAGENOMICS ANALYSIS	68
Abstract	68
1. Introduction.....	69
2. Materials and Methods	72
2.1. Sample Collection:	72
2.2. Sample Processing:.....	72
2.3. Nucleic acid Extraction:	74
2.4. Metagenomic analyses:.....	74
2.4.1. MG-RAST analyses:.....	75
2.4.2. METAVIR 2 analyses:.....	75
3. Results and Discussion:.....	76
3.1. Metagenome sequencing and assembly results:	76
3.2. Microbial diversity in WWTP:	78
3.2.1. Virus diversity in WWTP:	79
3.2.2. Bacterial diversity in WWTP:.....	80
3.3. Potential Pathogens in WWTP:	81
4. Conclusions:	84
5. Acknowledgement:.....	85
APPENDIX	86
REFERENCES.....	103
CHAPTER 5	110
VIRUS AND PHAGE DIVERSITY IN ACTIVATED SLUDGE FROM A CONVENTIONAL AND A MBR WASTEWATER UTILITY USING METAGENOMIC ANALYSES	110
Abstract	110
1. Introduction.....	110
2. Materials and Methods	112
2.1. Sample Collection:	112
2.2. Sample Processing:.....	112
2.2.1. Virus elution/isolation process:	113
2.2.2. Phage isolation:	113
2.3. Nucleic acid Extraction:	114
2.4. Metagenomic sequencing and analyses:.....	114
2.5. MG-RAST analyses:.....	115
3. Results and Discussion:.....	116
3.1. Metagenomic statistics:	116

3.2. Virus and phage diversity in sludge:	118
4. Conclusions:	121
5. Acknowledgement:.....	122
APPENDIX.....	123
REFERENCES.....	139
CHAPTER 6.....	143
CONCLUSIONS AND SIGNIFICANCE	143

LIST OF TABLES

Table 2.1: Metagenome analysis statistics (generated by MGRAST).....	24
Table 2.2: Presence of Phage lineages using BLAST searches.....	25
Table 3.1: Wastewater Treatment Characteristics.....	53
Table 3.2a: Metagenome analysis statistics in bacterial samples.....	54
Table 3.2b: Functional category Hit distribution	55
Table 3.3: Resistance groups for bacterial samples (MGRAST).....	60
Table 4.1: Microbial Contaminant Candidate List from EPA.....	87
Table 4.2: Virus and bacterial Pathogens detected in wastewater effluent.....	88
Table 4.3: Virus and bacterial Pathogens detected in raw and sludge samples	90
Table 4.4: Wastewater Treatment Plant Characteristics.....	92
Table 4.5: Metagenome analysis statistics for virus samples (by MetaVir).....	92
Table 4.6a: Metagenome analysis statistics for bacterial samples (generated by MGRAST).....	93
Table 4.6b: Functional category Hit distribution (Bacterial samples).....	93
Table 4.7: Taxonomic comparison heat map based contigs best BLAST hit ratios (number of hits for the genome divided by total number of hits in the metagenome).....	97
Table 4.8: Organism Abundance (Bacteria Phylum Distribution).....	99
Table 5.1: Characteristics of WWTPs.....	125
Table 5.2: Metagenome analysis statistics for sludge samples (generated by MGRAST).....	125
Table 5.3: Virus abundance in ELWWTP and TCWWTP activated sludge samples.....	130
Table 5.4: Phage abundance in ELWWTP and TCWWTP activated sludge samples.....	132

LIST OF FIGURES

Figure 2.1: (a) Subsystem functional barchart, (b) Functional distribution of “Virulence, Disease and Defense” subsystem, (c) Functional distribution of “Resistance to antibiotics”	27
Figure 2.2: Organism (genus) Tree.....	29
Figure 2.3: Concentration (copies/100mL) of (a) tetracycline resistant gene (Tet W), and (b) sulfonamide resistant gene (Sul I) abundance in Phage DNA from sludge samples..	30
Figure 3.1: Map showing location of sampling and sampling schematic.....	53
Figure 3.2: Schematic flowchart showing the procedure and methodology.....	54
Figure 3.3: Schematic showing the analysis step for the functional abundance in MGRAST....	55
Figure 3.4: Bacterial resistance to antibiotics and metals in activated sludge for East Lansing and Traverse City wastewater utilities.....	56
Figure 3.5: Bacterial resistance to antibiotics and metals in before disinfection water samples for East Lansing and Traverse City wastewater utility.....	57
Figure 3.6: Bacterial resistance to antibiotics and metals in after disinfection water samples for East Lansing and Traverse City wastewater utility.....	58
Figure 3.7: Relative abundance of resistance to antibiotics and metal compounds obtained from MGRAST.....	59
Figure 4.1: Location of two wastewater treatment plants selected for effluent sampling for investigation of the viral and bacterial community.....	94
Figure 4.2: Schematic flowchart showing the procedure followed for metagenomic analysis...	95
Figure 4.3: Rarefaction curve of species richness in virus and bacterial DNA enriched samples from effluents of two different WWTPs.....	96
Figure 4.4: Potentially pathogenic virus abundance in effluent samples.....	101
Figure 4.5: Potentially pathogenic bacteria diversity abundance in effluent samples.....	102
Figure 5.1: Schematic of methodology and location of sampling.....	124
Figure 5.2: Metagenome Summary in phage and virus DNA enriched samples from sludge of two different WWTPs.....	126

Figure 5.3: Rarefaction curve of species richness in phage and virus DNA enriched samples from sludge of two different WWTPs.....127

Figure 5.4: Virus diversity/organism abundance (family-level) based on best hit classification.....128

Figure 5.5: Phage diversity/organism abundance based on best hit classification.....129

CHAPTER 1

INTRODUCTION

Wastewater presents a time dynamic collection spot where many substances of physical, chemical and biological nature are brought together at one point (Sinclair et al. 2008). Wastewater treatment varies from one plant to another, however in general it is a multi-stage process that treats wastewater before it is discharged to a body of water, applied to land or re-used (Shannon et al. 2007). The efficiency of the wastewater treatment process depends on several factors like the type of biological treatment (for example conventional activated sludge process or MBRs (Membrane Bioreactor)), hydraulic residence time and solids retention time (Ma et al. 2011, Saikaly et al. 2005, Tchobanoglous et al. 2003). Changes in the bacterial community were also observed in response to changes in operational parameters of activated-sludge systems (Saikaly et al. 2005).

Studies have shown that potentially pathogenic bacteria were detected in the activated sludge and effluents from WWTPs (Ye and Zhang, 2011, Odjadjare, 2010). Conventional utilities and even state of the art WWTPs such as MBRs have been proven to release pathogenic viruses in the environment (Simmons and Xagorarakis, 2011, Bibby and Peccia, 2013). According to a recent study, human enteric viruses were detected in effluent from two different WWTPs using RT-QPCR (Kitajima et al. 2014). Another study based on characterizing effluent water quality from satellite MBRs facilities reported that adenoviruses were detected in effluent from all nine MBR facilities sampled (Hirani et al. 2013). There have been reports of finding pathogens in the effluent from different WWTPs even after disinfection treatment (Kitajima et al. 2014, Hirani et al. 2013, Simmon et al. 2011, Fong et al. 2010, Okoh et al. 2007).

Additionally, Antibiotic resistant bacteria and genes encoding antibiotic resistance are commonly detected at high rates and concentrations in wastewater samples (Munir et al. 2011, Zhang and Zhang 2011, Borjesson et al. 2009, Zhang et al. 2009a, Auerbach et al. 2007, Brooks et al. 2007, Kim and Aga 2007, Pruden et al. 2006, Reinthaler et al. 2003). Large numbers of antibiotic resistant organisms can survive in sewage and reach the wastewater treatment plant (Reinthaler et al. 2003, Guardabassi et al. 2002). Wastewater treatment plants (WWTPs) can be considered as an important reservoirs for the spread of antibiotic resistance to opportunistic pathogens and can stimulate horizontal gene transfer among microbial species.

Occurrence of antibiotic resistance bacteria (ARB) and antibiotic resistant genes (ARG) in our environment is a growing global health problem. Due to increasing evidences of antibiotic resistance in pathogenic and benign bacteria in our environment, an emerging threat to public and environmental health has been reported (Munir and Xagorarakis 2011, Knapp et al. 2009, Blasco et al. 2008). The use of numerous antimicrobial agents as treatments in animal, human, and plant health maintenance, is a worldwide practice providing both desirable and undesirable consequences (Munir et al. 2011). Links have been found to exist between antibiotic use and the emergence of antibiotic resistant genes (Gao et al. 2012). Studies have proven increase in antibiotic resistance strains that belong to pathogenic bacteria (Blasco et al., 2008, Peak et al., 2007).

Horizontal gene transfer in bacteria is an important process in accelerating the dispersal of ARGs in the environment (Colomer-Lluch et al. 2011a, Baquero et al. 2008, Sander and Schmieger 2001). Until the 1950s, when antibiotic resistance emerged worldwide, the significance of horizontal gene transfer for bacterial evolution was not recognized (Ochman et al. 2000). Horizontal gene transfer is the movement of genetic material among bacterial species

without cell division. In recent years, efforts have been made to study various gene transfer mechanisms involved in the spread of antibiotic resistance. Transformation is the direct uptake of naked DNA from the surroundings. It is the most common and widespread means of horizontal gene transfer. Conjugation is the transfer of DNA mediated by a conjugative or mobilizable genetic element (plasmids or transposons). It requires cell to cell contact and long fragments of DNA can be transferred through this mechanism. The transfer of DNA mediated by bacteriophage is known as transduction.

Bacteriophages play a major role in bacterial evolution facilitated by transferring virulence and antibiotic resistant genes to new bacterial hosts via the process of transduction (Mazaheri Nezhad Fard et al. 2011, Canchaya et al. 2004, Boyd and Brüssow 2002, Weinbauer and Rassoulzadegan 2003). Bacteriophages, also called phages, are viruses that infect bacteria. They all contain nucleic acid surrounded by a protein coat that makes them stick on to bacterial cell walls. When attached, they inject the DNA into the bacteria. Only few studies have been conducted to determine antibiotic resistant genes present in bacteriophage isolated from wastewater environments (Parsley et al. 2010, Colomer-Lluch et al. 2011a, Mazaheri Nezhad Fard et al. 2011, Prescott 2004, Muniesa et al. 2004a). Recently the role of phages in the spread of ARGs in the environment has been studied (Colomer-Lluch et al. 2011). That study highlights the potential role of phages in the spread of β lactamase genes in urban sewage and river water samples and found that phages are a suitable candidate to act as reservoir for the spread of ARGs in the environment. Another study was done on enterococcal bacteriophages which have been shown to play a role in successful transfer of antibiotic resistant genes as tetracycline (tetM) and gentamicin (ant2-I) resistance between the same and different enterococcal species (Mazaheri Nezhad Fard et al. 2011).

Wastewater microbial diversity including potential pathogens and antibiotic resistant bacteria is vast and still not clearly characterized. Molecular biology is currently being revolutionized with the emergence of next generation technology. It has been reported that these next-generation DNA sequencing methods have the ability to significantly help to accelerate biological research (Shendure and Ji 2008). A field known as metagenomics is fast evolving and provides a way of characterizing the entire microbial communities (microbiome). Environmental metagenomics is the study of organisms in a microbial community based on analyzing the DNA within an environmental sample. Environmental metagenomics as a field was extremely limited prior to the advent of next-generation sequencing (NGS). Next-generation sequencing has substantially widened the scope of metagenomic analysis of environmentally derived samples (Mardis 2008). The high demand for low-cost sequencing methods has motivated the improvement of sequencing technologies. The advent of NGS has decrease the time required and the cost of complete genome sequencing (Subramanian et al. 2010, Mardis 2008, Soni and Meller 2007, Shendure and Ji 2008). These recent technologies allow us to sequence DNA and RNA much more quickly and cheaply than the previously used Sanger sequencing. NGS provides researchers the capability to profile entire microbial communities from complex samples, discover new organisms, and explore the dynamic nature of microbial populations under changing conditions.

Metagenomic technologies present an opportunity for generating an improved understanding of the water microbiome and thus enhancing microbial water quality and water safety (Aw and Rose, 2013; Edwards and Rohwer, 2005). Various methods have been applied to investigate microbial community in wastewater but they provide only limited information compared to latest emerging high throughput sequencing technologies. According to Zhang et al

(2011), a comprehensive characterization of the vast microbial community present in activated sludge systems is hindered by the low sequencing depth of the traditional PCR-cloning approach (Zhang et al. 2011). Next-generation DNA sequencing has recently been applied to study viral metagenomes in different environmental samples (Alhamlan et al. 2013, Gomez-Alvarez et al. 2012, Hu et al. 2012, Bibby et al. 2011, Wommack et al. 2011, Tamaki et al. 2011, Rosario et al. 2009). With the help of metagenomic tools, microbial communities related with wastewater systems could easily be analyzed. The objective of this study is to provide metagenomic insights into bacterial, viral and phage diversity and resistance to antibiotics and metal compounds in wastewater utilities.

REFERENCES

REFERENCES

1. Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R., & Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *Journal of microbiological methods*, 92(2), 183-188.
2. Auerbach, E.A., Seyfried, E.E. and McMahon, K.D. (2007) Tetracycline resistance genes in activated sludge wastewater treatment plants. *Water research* 41(5), 1143-1151.
3. Aw, T. G., & Rose, J. B. (2013). Get Ready for The Future: Viral Metagenomics for Defining Healthy Water. *Proceedings of the Water Environment Federation*, 2013(10), 5098-5102.
4. Bibby, K., Viau, E., & Peccia, J. (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Letters in applied microbiology*, 52(4), 386-392.
5. Blasco, M.D., Esteve, C. and Alcaide, E. (2008). Multiresistant waterborne pathogens isolated from water reservoirs and cooling systems. *Journal of Applied Microbiology* 105(2), 469-475.
6. Borjesson, S., Melin, S., Matussek, A. and Lindgren, P.E. (2009). A seasonal study of the *mecA* gene and *Staphylococcus aureus* including methicillin-resistant *S. aureus* in a municipal wastewater treatment plant. *Water Res* 43(4), 925-932.
7. Brooks, J.B.J., Maxwell, S.M.S., Rensing, C.R.C., Gerba, C.G.C. and Pepper, I.P.I. (2007). Occurrence of antibiotic-resistant bacteria and endotoxin associated with the land application of biosolids. *Can J Microbiol* 53(5), 616-622.
8. Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nature Reviews Microbiology*, 3(6), 504-510.
9. Gomez-Alvarez, V., Revetta, R. P., & Santo Domingo, J. W. (2012). Metagenomic analyses of drinking water receiving different disinfection treatments. *Applied and environmental microbiology*, 78(17), 6095-6102.
10. Guardabassi, L., Lo Fo Wong, D. and Dalsgaard, A. (2002). The effects of tertiary wastewater treatment on the prevalence of antimicrobial resistant bacteria. *Water research* 36(8), 1955-1964.
11. Hu, M., Wang, X., Wen, X., & Xia, Y. (2012). Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresource technology*, 117, 72-79.

12. Kim, S. and Aga, D.S. (2007). Potential ecological and human health impacts of antibiotics and antibiotic-resistant bacteria from wastewater treatment plants. *Journal of Toxicology and Environmental Health, Part B* 10(8), 559-573.
13. Knapp, C.W., Dolfing, J., Ehlert, P.A. and Graham, D.W. (2009). Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environmental science & technology* 44(2), 580-587.
14. Ma, Y., Wilson, C.A., Novak, J.T., Riffat, R., Aynur, S., Murthy, S. and Pruden, A. (2011). Effect of various sludge digestion conditions on sulfonamide, macrolide, and tetracycline resistance genes and class I integrons. *Environ Sci Technol* 45(18), 7855-7861.
15. Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402.
16. Munir, M. and Xagorarakis, I. (2011). Levels of Antibiotic Resistance Genes in Manure, Biosolids, and Fertilized Soil. *Journal of Environment Quality* 40(1), 248.
17. Munir, M., Wong, K. and Xagorarakis, I. (2011). Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. *Water Res* 45(2), 681-693.
18. Peak, N., Knapp, C.W., Yang, R.K., Hanfelt, M.M., Smith, M.S., Aga, D.S. and Graham, D.W. (2007). Abundance of six tetracycline resistance genes in wastewater lagoons at cattle feedlots with different antibiotic use strategies. *Environmental microbiology* 9(1), 143-151.
19. Pruden, A., Pei, R., Storteboom, H. and Carlson, K.H. (2006). Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environmental science & technology* 40(23), 7445-7450.
20. Reinthaler, F., Posch, J., Feierl, G., Wüst, G., Haas, D., Ruckebauer, G., Mascher, F. and Marth, E. (2003). Antibiotic resistance of *E. coli* in sewage and sludge. *Water research* 37(8), 1685-1690.
21. Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., & Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental microbiology*, 11(11), 2806-2820.
22. Saikaly, P.E., Stroot, P.G. and Oerther, D.B. (2005). Use of 16S rRNA gene terminal restriction fragment analysis to assess the impact of solids retention time on the bacterial diversity of activated sludge. *Appl Environ Microbiol* 71(10), 5814-5822.
23. Shannon, K. E., Lee, D. Y., Trevors, J. T., and Beaudette, L. A. (2007). Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Science of the total environment*, 382(1), 121-129.

24. Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology* 26(10), 1135-1145.
25. Sinclair, R. G., Choi, C. Y., Riley, M. R., and Gerba, C. P. (2008). Pathogen surveillance through monitoring of sewer systems. *Advances in applied microbiology*, 65, 249.
26. Soni, G.V. and Meller, A. (2007). Progress toward ultrafast DNA sequencing using solid-state nanopores. *Clinical chemistry* 53(11), 1996-2001.
27. Subramanian, S., Huynen, L., Millar, C.D. and Lambert, D.M. (2010). Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. *BMC Evol Biol* 10, 387.
28. Tamaki, H., Zhang, R., Angly, F. E., Nakamura, S., Hong, P. Y., Yasunaga, T., ... & Liu, W. T. (2012). Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environmental microbiology*, 14(2), 441-452.
29. Tchobanoglous, G., Burton, F.L. and Stensel, H.D. (2003). *Solution Manual for Use with Wastewater Engineering: Treatment and Reuse*, McGraw-Hill.
30. Wommack, K. E., Srinivasiah, S., Liles, M., Bhavsar, J., Bench, S., Williamson, K. E., & Polson, S. W. (2011). Metagenomic contrasts of viruses in soil and aquatic environments. *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*. New York: John E. Wiley & Sons.
31. 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), e26041.
32. Zhang, X.-X. and Zhang, T. (2011). Occurrence, abundance, and diversity of tetracycline resistance genes in 15 sewage treatment plants across China and other global locations. *Environmental science & technology* 45(7), 2598-2604.
33. Zhang, Y., Marrs, C.F., Simon, C. and Xi, C. (2009a). Wastewater treatment contributes to selective increase of antibiotic resistance among *Acinetobacter* spp. *Science of the Total Environment* 407(12), 3702-3706.

CHAPTER 2

PHAGE AND ANTIBIOTIC RESISTANCE GENES IN A CONVENTIONAL WASTEWATER TREATMENT PLANT

Munir M., T. Marsh, and I. Xagorarakis. *Submitted for consideration to Water Research.*

Abstract

Wastewater treatment plants (WWTPs) can be considered as an important reservoirs for the spread of antibiotic resistance to opportunistic pathogens and can stimulate horizontal gene transfer among microbial species. Bacteriophages exist in most environments and may play a major role in the dissemination of antibiotic resistant genes (ARGs) within WWTPs. Phage diversity was studied by next generation sequencing on sludge samples (before and after DNase treatment) with Illumina (Miseq). Sludge samples were collected from a conventional WWTP in Michigan. A method for phage DNA isolation was optimized using PEG (polyethylene glycol) precipitation and DNase (deoxyribonuclease) treatment. Metagenome data analysis revealed that after DNase treatment and assembly of contigs, the activated returned sludge (RAS) sample contained 21,985 contigs totaling 17,227,533 basepairs with an average length of 783 bps and primary sludge (PS) contained 2,870 contigs sequences totaling 2,292,422 basepairs with an average length of 798 bps. On a genus level, *Burkholderia phage*, *Coliphage*, *Enterobacteria phage*, and *Pseudomonas phage* are present in all the samples. Phages infecting *Burkholderia cepacia*, *Edwardsiella*, *Mycobacterium*, *Salmonella*, *Vibrioe* and *Xanthomonas citri* were detected only in RAS samples while phages infecting *Bacillus*, *Brochothrix*, *Lactobacillus*, *Listeria*, *Phormidium* and *Staphylococcus* were found only in PS samples. Additionally, phage DNA was isolated and screened for ARGs (tetracycline resistant genes (Tet-W and Tet-O) and sulfonamide resistant gene (Sul-I)) using real-time Q-PCR. We have detected ARGs in phage

DNA with concentrations ranging from 3.84×10^2 - 8.14×10^3 copies/100mL for Tet-W gene and 5.89×10^4 - 7.9×10^4 copies/100mL for Sul-I gene. In addition, phage metagenomes were searched for functional signatures of resistance genes. Metagenomics analysis revealed that most of the antibiotic resistance belongs to methicillin, fluoroquinolones and beta-lactamase group of antibiotics. This work presents the diversity and occurrence of phages in sludge samples and indicates that there is a substantial shift in the phage community over the course of the activated sludge process, thus suggesting that within the activated sludge the phage populations are dynamic. This work indicates that phage DNA was associated with antibiotic resistant genes in wastewater.

Keywords: wastewater, activated sludge, antibiotic resistant gene, next-generation sequencing, Bacteriophage metagenome

1. Introduction

Viruses are the most abundant and most diverse group of biological entities. Bacteriophage, viruses that attack bacteria (hereafter referred to as phage), have abundance and distribution that in most cases reflects that of their host organisms. Contemporary investigations focusing on the ecology and genetics of phage take advantage of metagenomics that can yield useful information based on the amount of coverage of particular phages/gene sets present in environmental samples (Clokier et al. 2011). Phages contain nucleic acid surrounded by a protein coat that makes them stick on to bacterial cell walls. When attached, they inject the DNA into bacteria where transcription, replication and assembly of new phage take place. Horizontal gene transfer in bacteria is an important process in accelerating the dispersal of ARGs in the environment (Colomer-Lluch et al. 2011a, Baquero et al. 2008, Sander and Schmieger 2001). Phages play a major role in bacterial evolution by transferring antibiotic resistant genes to new

bacterial hosts via the process of transduction, which is one of the mechanism of horizontal gene transfer (Muniesa et al. 2013a,b, Mazaheri Nezhad Fard et al. 2011, Canchaya et al. 2004, Muniesa et al. 2004, Boyd and Brüssow 2002, Weinbauer and Rassoulzadegan 2003).

Due to the increasing evidence of antibiotic resistance in pathogenic and benign bacteria in our environment, an emerging threat to public and environmental health has been postulated (Munir and Xagorarakis 2011, Knapp et al. 2009, Blasco et al. 2008). Antibiotic resistant bacteria and genes encoding antibiotic resistance are commonly detected at high rates and concentrations in wastewater samples (Munir et al. 2011, Zhang and Zhang 2011, Borjesson et al. 2009, Zhang et al. 2009, Auerbach et al. 2007, Brooks et al. 2007, Kim and Aga 2007, Pruden et al. 2006, Reinthaler et al. 2003). Large numbers of antibiotic resistant organisms can survive in sewage and reach the wastewater treatment plant (Reinthaler et al. 2003, Guardabassi et al. 2002). A recent study suggested that multidrug resistant genes even survive through several wastewater treatment units, including disinfection (Luo et al. 2014). A strong link has been reported between wastewater and antibiotic resistance (Börjesson S et al. 2009, Volkmann et al. 2004, and Schwartz et al. 2003).

Our understanding of the role that phage play in the dissemination of antibiotic resistances is at an early stage with only a few studies addressing this (Parsley et al. 2010, Colomer-Lluch et al. 2011a, Mazaheri Nezhad Fard et al. 2011, Prescott 2004, Muniesa et al. 2004). According to recent literature review, ARGs related with phages have been identified in different environmental samples. For example phages have been termed a reservoir for the spread of β lactamase genes in urban sewage and river water samples (Colomer-Lluch et al. 2011a). Gene resistant to β -lactam antibiotics have also been identified in fecal waste from cattle, pigs and poultry using PCR and QPCR (Colomer-Lluch et al. 2011b). A group of antibiotic

resistance genes were detected in bacteriophage DNA isolated from human fecal samples (Quiros et al. 2014). Thus, bacteriophages have been regarded as potential vector for ARG transfer. Wastewater can provide favorable conditions for the growth and propagation of antibiotic resistant bacteria and their genes. Large amounts of antibiotics are released into municipal wastewater due to incomplete metabolism in humans or to disposal of unused antibiotics. Once in the wastewater stream they can exert selective pressure for or maintain resistance among microorganisms (Allen et al. 2010, Nagulapally et al. 2009). Activated sludge has been referred to be a “hot-bed” for horizontal gene transfer and selection of antibiotic resistant genes among aquatic bacteria (Guardabassi et al. 2002). The design of activated sludge in WWTPs is primarily focused to maximize biological substrate removal by promoting factors for nurturing bacterial retention and growth (Kim et al. 2010) and promoting cellular interactions among diverse microorganisms. Thus, it provides great potential for the lateral transfer of ARGs between microbes in activated sludge (Parsley et al. 2010) and is characterized with high concentration of microbial community that facilitates horizontal gene transfer (HGT) of ARG via mobile genetic elements (Zhang and Zhang 2011).

The objective of this study was to describe and compare the diversity of phages present in primary sludge and returned activated sludge using metagenomic investigations. Further, the goal was to detect ARGs in bacteriophage in order to assess the likely occurrence of transductional transfer within wastewater treatment plants.

2. Materials and Methods

2.1. Sample Collection: Returned activated sludge (RAS) and primary sludge (PS) samples were collected from East Lansing WWTP in Michigan (U.S.A.) in 2012. Samples were kept on ice

and were transported to the Laboratory at Michigan State University (East Lansing, U.S.A.) for immediate processing.

2.2. Sample Processing: Samples were induced with Mitomycin C (1µg/ml) and incubated at room temperature for 24 hrs while gently shaking (150 rpm). Several drops of chloroform were added to the samples to complete lysis and incubation was continued for another 15-30 mins. The samples (250-300mL of sludge) were then centrifuged at 3396 xg for 45 minutes in F10S-6X500Y rotor and the supernatant was carefully decanted. The pellet was saved for bacterial DNA extraction. Each supernatant was filtered through a 0.22 µm filter (Millipore, Billerica, MA) and then the bacteriophages were precipitated with PEG-NaCl (Colomer-Lluch et al. 2011a, Sander and Schmieger 2001, Muniesa et al. 2004). The PEG precipitate was collected by centrifugation at 10000xg in an aerosol-tight fixed-angle rotor at 4°C for 40 minutes. The supernatant was carefully decanted and the pellet was resuspended in 1.0 ml of SM buffer (Yamamoto et al., 1970). Any free DNA that co-purified with bacteriophage was removed by digestion with DNase I (100 Units/mL) (Colomer-Lluch et al. 2011a) in half of each of the samples. The DNase-treated phage preparations and the ones without DNase were stored at -80°C until DNA extraction was performed for molecular analysis. The volume of all the samples initially collected for processing was taken into account when calculating final concentrations.

A positive control test was conducted initially using Coliphage T4 and E.coli BRE607 on a M9 supplement media in a Plaque Assay. A high titre of phages (10^8 - 10^{10} pfu/mL) was grown. The phage was tested after phage isolation by PEG precipitation using plaque assay to confirm the method.

2.3. DNA Extraction: Phage DNA and bacterial DNA was extracted (from samples with and without DNase treatment) using a MagNA Pure Compact DNA extractor (Roche Applied Science, Indianapolis, IN, USA) following the protocol in the manufacturer's manual. The MagNA Pure Compact utilizes a magnetic-bead technology for the isolation process. Sample amount of 400 μ L was loaded in the system and the elution volume was 100 μ L. The extracts were stored in a freezer at -20°C.

2.4. Quantification: Antibiotic resistance genes were quantified in these samples using previously developed assays in our lab (Munir et al. 2011). In these samples, tetracycline resistance genes (*tetO* and *tetW*) and sulfonamide resistance gene (*sulI*) were detected using real-time qPCR with SYBR Green method, which was optimized using previously described primers (Aminov et al. 2001, Pei et al. 2006). All the qPCR reactions were performed with a Roche LightCycler 1.5 (Roche Applied Science, Indianapolis, IN, USA). All reactions were done in triplicate.

2.5. Metagenomic analyses: Bacteriophage-enriched DNA isolated from the sludge samples was sequenced on an Illumina platform (Illumina MiSeq, Roche Technologies) at The Research Technology Support Facility (RTSF) genomic center at Michigan State University generating 150 bp paired-end reads. The sequences were assembled using an integrated pipeline for de NOVO assembly of microbial genomes. An assembly pipeline called A5 (<http://ged.msu.edu/angus/2013-04-assembly-workshop/assembly-with-a5.html>) was applied that simplifies the entire genome assembly process by automating sequence data cleaning, error correction, assembly, and quality control and automated assembly parameter selection (Tritt et al. 2012).

2.5.1. Blast analyses: Standard nucleotide BLAST analyses were conducted against the NCBI non-redundant nucleotide sequence (nr/nt) database to identify members of gene families. Blast (<http://www.ncbi.nlm.nih.gov/blast>) analysis was performed on the contigs file generated from the A5 assembler to determine the phage metagenome of the activated sludge.

2.5.2. MG-RAST analyses: The assembled data from the two samples were also analyzed using MG-RAST 3.3.1 (Meyer et al. 2008). Data was analyzed based on organism abundance and on the functional distribution at the subsystem hierarchy with maximum E- value cutoff of $1E^{-5}$, minimum percent identity cutoff of 60% and minimum alignment length cutoff of 15bps. The displayed data has been normalized to values between 0 and 1 to allow for comparison of differently sized samples. Each of the categories was further studied for detailed analysis further exploring each category in more detail.

3. Results and Discussion:

3.1. Metagenome sequencing and assembly results:

In order to investigate the phage community in activated sludge, next-generation sequencing was used. Bacteriophage-enriched DNA isolated from the sludge samples at two different locations within the wastewater treatment plant were sequenced using an Illumina Miseq and assembled using A5 pipeline. Both the DNase-treated and non- treated phage preparations were used in metagenomics analyses to observe the activity of DNase on the samples in detecting free DNA that co-purified with bacteriophage during the phage isolation process. Metagenome analysis revealed that after DNase treatment the returned activated sludge (RAS) sample contained 21,985 sequences totaling 17,227,533 basepairs (bps) with an average length of 783 bps and the primary sludge (PS) sample contained 2,870 contigs sequences totaling 2,292,422 bps with an

average length of 798 bps. The RAS samples without the DNase treatment contained 23,663 sequences totaling 18,332,554 bps with an average length of 774 bps and PS samples without DNase treatment contained 7,796 sequences totaling 5,511,578 bps with an average length of 706 bps. The decline in the number of contigs in each set of samples indicates that some free DNA was lost as a result of DNase treatment indicating that there was some extracellular DNA present in the PEG precipitation process for phage isolation. Analysis statistics for all the samples are shown in Table 2.1.

Annotation of all the reads for the functional distribution at the subsystem hierarchy showed that 63.93 % and 46.56% of sequences were predicted as “Phages or, Prophages” in the metagenome from RAS and PS samples after DNase treatment respectively. Figure 2.1 illustrates the distribution of functional categories at the highest level supported by subsystems analysis. The presence of phages is indicated in the bar chart with phages occupying the majority on 0 to 1 scale, in both RAS and PS samples (Figure 2.1a). In this figure, the membrane transport functional category, along with cell division and cell cycle function showed lesser values after DNase treatment in both in RAS and PS samples suggesting that there was loss of free DNA after DNase treatment. The data also demonstrate the occurrence of virulence, disease and defense factors in the sludge samples occupying 10-38% of functional hits. Analysis of just the ‘virulence, disease and defense’ functional category, revealed a higher resistance to antibiotics in PS samples compared to RAS sample (Figure 2.1). Deeper analysis of the metagenomic data revealed that most of the antibiotic resistance belonged to methicillin, fluoroquinolones and beta-lactamase group of antibiotics (Figure 2.1c).

3.2. Phage Diversity:

The diversity of phages present in RAS and PS was studied using MGRAST v3.0 pipeline. The MG-RAST pipeline analysis includes the phylogenetic comparisons and functional annotations against the database. Standard nucleotide BLAST analyses were also conducted against the NCBI non-redundant nucleotide sequence (nr/nt) database to identify members of gene families. Figure 2.2 shows the phage diversity present in RAS and PS samples on a genus level classification. The stacked bar chart indicates the abundance of each genus in each of the samples analyzed. Based on MGRAST analysis, all the samples showed presence of *Chlorovirus*, *Microvirus*, *Siphoviridae*, *Lambda-like viruses*, *Podoviridae*, *P22-like viruses*, *T4-like viruses*, *SPO1-like viruses* and *Myoviridae* (Figure2). The detailed list of types of phages present in all the samples in this study is presented in Table 2.2 (Blast search analysis). *Burkholderia phage*, *coliphage*, *Enterobacteria phage*, and *Pseudomonas phage* were present in both RAS and PS sample before and after treatment. *Burkholderia cepacia phage*, *Edwardsiella phage*, *Mycobacterium phage*, *Salmonella phage*, *Vibrio phage* and *Xanthomonas citri phage* were detected only in returned activated sludge (RAS) samples. *Bacillus phage*, *Brochothrix phage*, *Lactobacillus phage*, *Listeria phage*, *Phormidium phage*, and *Staphylococcus phage* were found only in primary sludge (PS) samples. The trend in the abundance of phages in sludge samples indicates that there was a substantial shift in the phage community over the course of the activated sludge process, indicating that within the activated sludge the phage populations are dynamic. Reasoning for some of the phages that are only detected in RAS but not in PS, this may be due to the fact that PS sample have generated significantly small sequence size compared to RAS samples with sequencing.

The phage diversity detected in this study using Miseq (Illumina) sequencing platform is similar to the diversity presented in the literature. *Myoviridae*, *Podoviridae*, *Siphoviridae*, are the most common group of phages found in wastewater samples (Colomer-Lluch et al. 2011). Parsley and his team followed shotgun library approach to study activated sludge samples finding *Myoviridae* (40.3%), *Siphoviridae* (31.9%), *Podoviridae* (25.6%) and considered unclassified phages (2.2%) (Parsley et al. 2010). *Myoviridae* and *Siphoviridae* were also detected when electron microscopy was used on sewage and river water sample (Colomer-Lluch et al. 2011). Pyrosequencing discovered similar results on dairy manure wastewater lagoons (Alhamlan et al. 2013).

3.3. Antibiotic Resistant Gene Diversity:

The phage metagenome was searched for functional signatures of resistance genes using MGRAST. Greater percentage of antibiotic resistant genes was observed in PS samples compared to RAS sample when analyzing the ‘virulence, disease and defense’ functional category (Figure 2.1b) according to MGRAST analysis. It was found that most of the antibiotic resistance genes conferred resistance to methicillin, fluoroquinolones and beta-lactamase group of antibiotics (Figure 2.1c). Further exploring the phage metagenome for ARGs in MGRAST, it was found that RAS sample with DNase treatment contain proteins for Oxetanocin resistance and Vancomycin resistance whereas RAS samples without DNase treatment contain proteins for multiple antibiotic resistance, Oxetanocin resistance, quaternary ammonium compound-resistance and Tellurium resistance. PS sample with DNase treatment contain proteins for only Oxetanocin resistance whereas no data was returned for abundance of ARGs in the metagenome for PS sample with DNase treatment An interesting copper resistance protein was also detected

in phage fraction obtained from RAS sample without DNase treatment. It has been suggested that presence of metals in wastewater treatment can also drive for selection of antibiotic resistance among bacteria (Peltier et al. 2010; Knapp et al. 2011; Wright et al. 2006). Wright et al. (2006) concluded that metal exposure can directly select for metal-resistance while co-selecting for antibiotic-resistant bacteria. Also, the presence of Staphylococcus phage indicated by BLAST analysis and the presence of methicillin resistance in Staphylococci from MGRAST analysis showed that the results are related (Figure 2.1).

3.4. Concentration of ARGs:

Antibiotic resistance genes were quantified with QPCR using previously developed assays in our lab (Munir et al. 2011). Phage isolate (with- and without-DNase treatment) and bacterial isolate from returned activated sludge and primary sludge samples collected from East Lansing WWTP in Michigan were tested for detectable tetracycline resistance genes (*tetO* and *tetW*) and sulfonamide resistance gene (*sulI*). Concentrations of ARGs in phage DNA with DNase treatment for RAS and PS samples were found to be 3.84×10^2 and 8.14×10^3 copies/100mL for Tet-W gene and 5.89×10^4 and 7.9×10^4 copies/100mL for Sul-I gene, respectively (Figure 3). Whereas, concentrations of ARGs in phage DNA of RAS and PS samples without DNase treatment was found to be 2.14×10^3 and 2.5×10^4 copies/100mL for Tet-W gene and 4.17×10^5 and 1.19×10^5 copies/100mL for Sul-I gene, respectively (Figure 2.3). Tet-O gene was not detected in these samples. Concentrations of ARGs in bacterial DNA of RAS and PS samples were found to be 1.48×10^7 and 1.33×10^9 copies/100mL for Tet-W gene and 1.63×10^9 and 1.55×10^9 copies/100mL for Sul-I gene respectively (Figure 2.3). The concentration of phage associated ARGs detected in phage DNA was much lower than that present in the

fraction of bacterial DNA based on the same volume comparison. There is about 4-5 log difference in the concentration of ARGs between the two fraction of DNA in each of RAS and PS samples. A remarkable difference was seen in phage DNA before and after DNase treatment indicating the presence of free DNA containing ARGs that have been digested by DNase.

Our study also detected oxetanocin and vancomycin resistance (MGRAS^T) along with sulfonamide resistant gene Sul-I. A recent study using shotgun library approach found that phages appear to carry partial genes that may be responsible for resistance to tetracycline, ampicillin, acriflavine, and bleomycin, few others in activated sludge sample (Parsley et al. 2010). Our work is consistent with repeated isolations of antibiotic resistant bacteria from wastewater treatment plants and the detection of resistance determinants using cultivation independent techniques. More work is needed to understand the importance of phages and their role in ARG transfer among bacterial community in wastewater treatment plants.

4. Conclusions:

- Phage diversity was studied by next generation sequencing on sludge samples (before and after DNase treatment) with Illumina (Miseq).
- On a genus level, Burkholderia phage, Coliphage, Enterobacteria phage, and Pseudomonas phage are present in all the samples. Burkholderia cepacia phage, Edwardsiella phage, Mycobacterium phage, Salmonella phage, Vibrio phage and Xanthomonas citri phage were detected only in RAS samples. Bacillus phage, Brochothrix phage, Lactobacillus phage, Listeria phage, Phormidium phage, Staphylococcus phage and Sugarcane mosaic virus were found only in PS samples.

- Concentration of ARGs detected in phage DNA in all samples ranges from 3.84×10^2 - 8.14×10^3 copies/100mL for Tet-W gene and 5.89×10^4 - 7.9×10^4 copies/100mL for Sul-I gene.
- This work presents the diversity of phages in sludge samples and indicates that phage DNA was associated with antibiotic resistant genes in wastewater.

5. Acknowledgement:

We would like to thank the manager of the East Lansing Wastewater Treatment Plant for providing the samples and information needed for this study. Also, we would like to acknowledge bioinformatic support and assistance provided by Bioinformatic team at High Performance Computing Center (HPCC) at Michigan State University and a special thank to Bioinformatic Research Specialist Dr. Tracy Teal.

APPENDIX

Table 2.1: Metagenome analysis statistics (generated by MGRAST).

	RAS w/ DNase	RAS w/o DNase	PS w/ DNase	PS w/o DNase
Raw bp Count	17,227,533 bp	18,332,554 bp	2,292,422 bp	5,511,578 bp
No. of contigs	21,985	23,663	2870	7,796
Mean Sequence Length	783 ± 1053 bp	774 ± 1017 bp	798 ± 926 bp	706 ± 696 bp
Artificial Duplicate Reads: Sequence Count	884	677	98	229
Post QC: bp Count	13,742,435 bp	14,561,037 bp	1,727,518 bp	4,452,805 bp
Post QC: No. of contigs	20,523	22,301	2650	7,287
Post QC: Mean Sequence Length	669 ± 456 bp	652 ± 445 bp	651 ± 439 bp	611 ± 364 bp

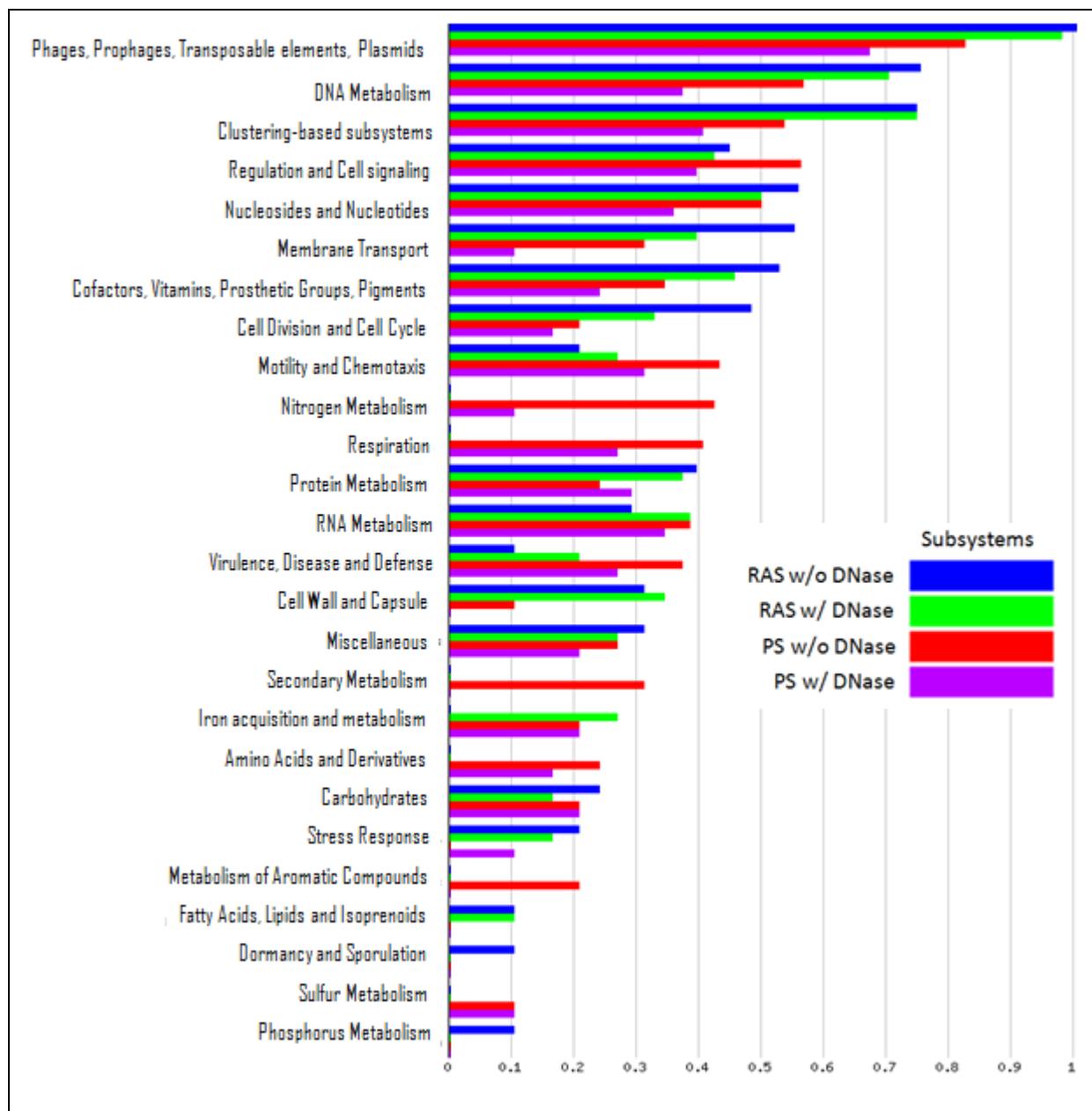
Note: Sequences were assembled using A5 pipeline assembly and contigs generated were analyzed on MGRAST; Abbreviation: bp= base pair

Table 2.2: Presence of Phage lineages using BLAST searches.

	RAS w/ DNase	RAS w/o DNase	PS w/ DNase	PS w/o DNase
Genus names				
Burkholderia phage	✓	✓	✓	✓
Coliphage	✓	✓	✓	✓
Enterobacteria phage	✓	✓	✓	✓
Pseudomonas phage	✓	✓	✓	✓
EBPR podovirus	✓	✓		✓
Burkholderia cepacia phage	✓	✓		
Edwardsiella phage	✓	✓		
Vibrio phage	✓	✓		
Xanthomonas citri phage	✓	✓		
Mycobacterium phage	✓	✓		
Salmonella phage	✓	✓		
Bacillus phage			✓	✓
Brochothrix phage			✓	✓
Lactobacillus phage			✓	✓
Listeria phage			✓	✓
Phormidium phage			✓	✓
Staphylococcus phage			✓	✓
Klebsiella phage	✓			✓
Environmental Halophage	✓			✓
Escherichia phage	✓			✓
Burkholderia cenocepacia phage	✓		✓	
Persicivirga phage		✓		✓
Helicobacter phage		✓		
Iodobacteriophage		✓		

Table 2.2 (cont'd)

Aeromonas phage		✓		
Bordetella phage		✓		
Caulobacter phage		✓		
Cronobacter phage	✓			
Erwinia amylovora phage		✓		
Enterobacter phage		✓		
Leptospira biflexa temperate bacteriophage		✓		
Pseudomonas aeruginosa phage		✓		
Rhizobium phage		✓		
Rhodobacter phage	✓			
Streptococcus phage	✓		✓	
Synechococcus phage	✓			
Thermus phage		✓		
Geobacillus virus			✓	
Acanthamoeba castellanii mamavirus				✓
Lactobacillus johnsonii prophage				✓
Lactobacillus plantarum bacteriophage			✓	
Listeria bacteriophage			✓	
Megavirus				✓
Rhodococcus phage			✓	

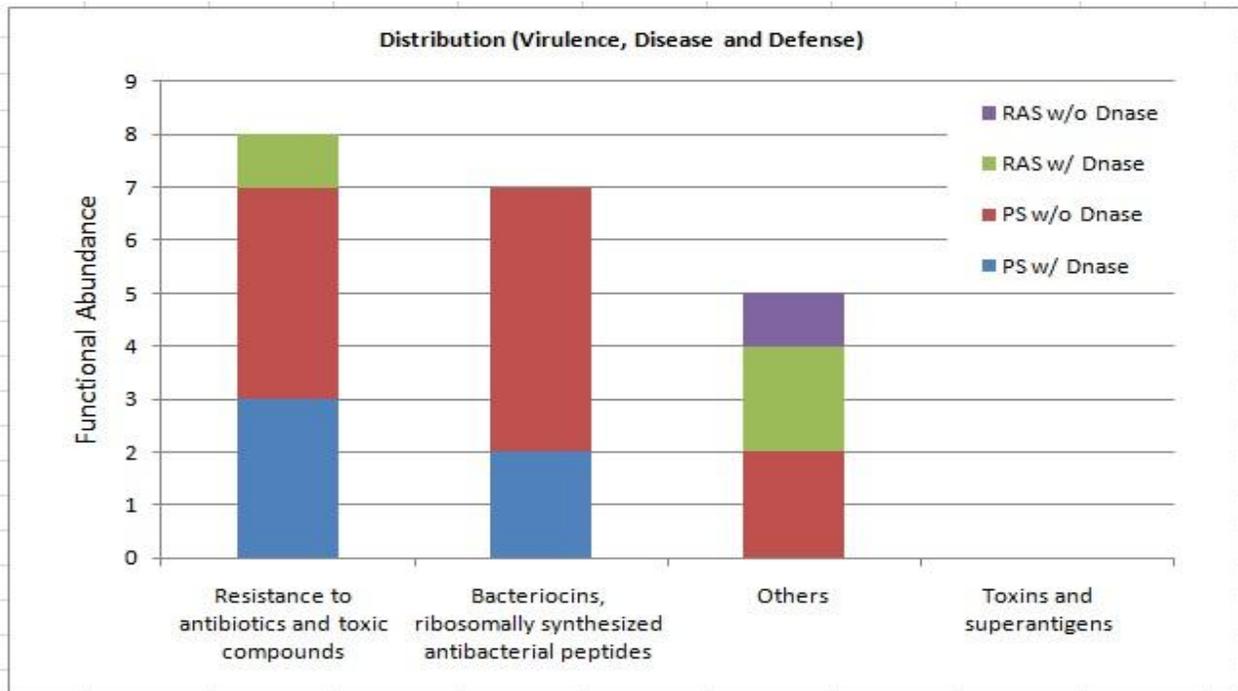


(a)

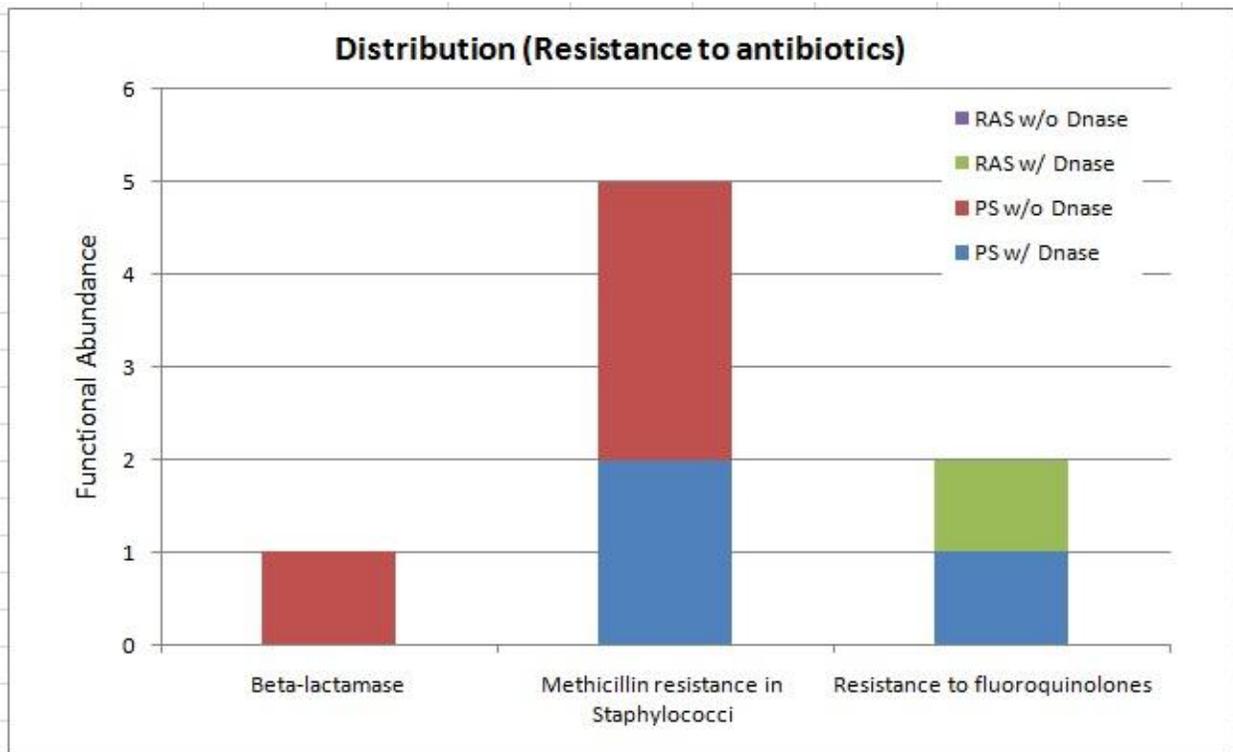
Figure 2.1: (a) Subsystem functional barchart, (b) Functional distribution of “Virulence, Disease and Defense” subsystem, (c) Functional distribution of “Resistance to antibiotics”.

Note: The data was compared to Subsystems using a maximum *e*-value of $1e-5$, a minimum identity of 60 %, and a minimum alignment length of 15 measured in aa for protein and bp for RNA databases.

Figure 2.1 (cont'd)



(b)



(c)

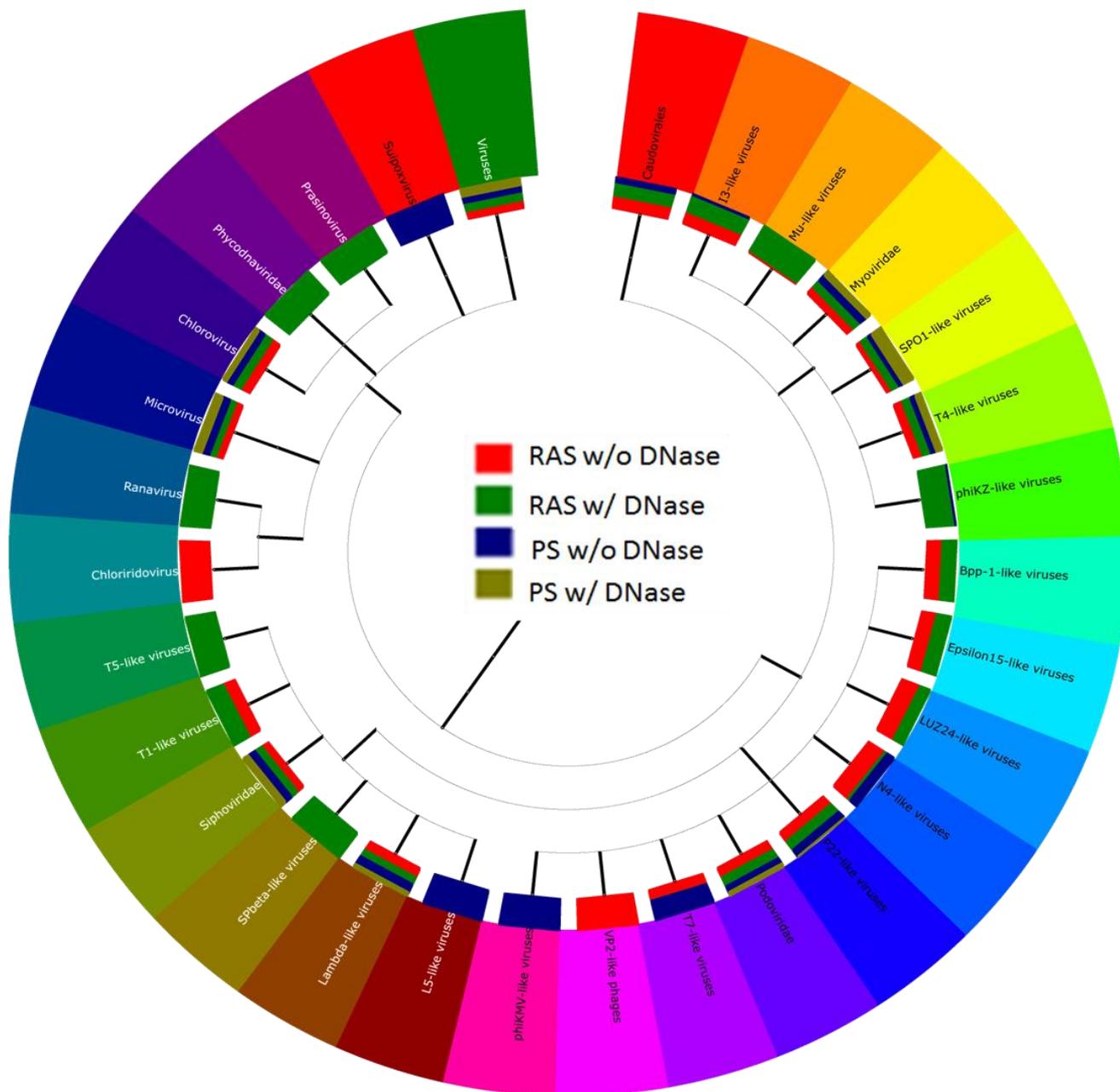
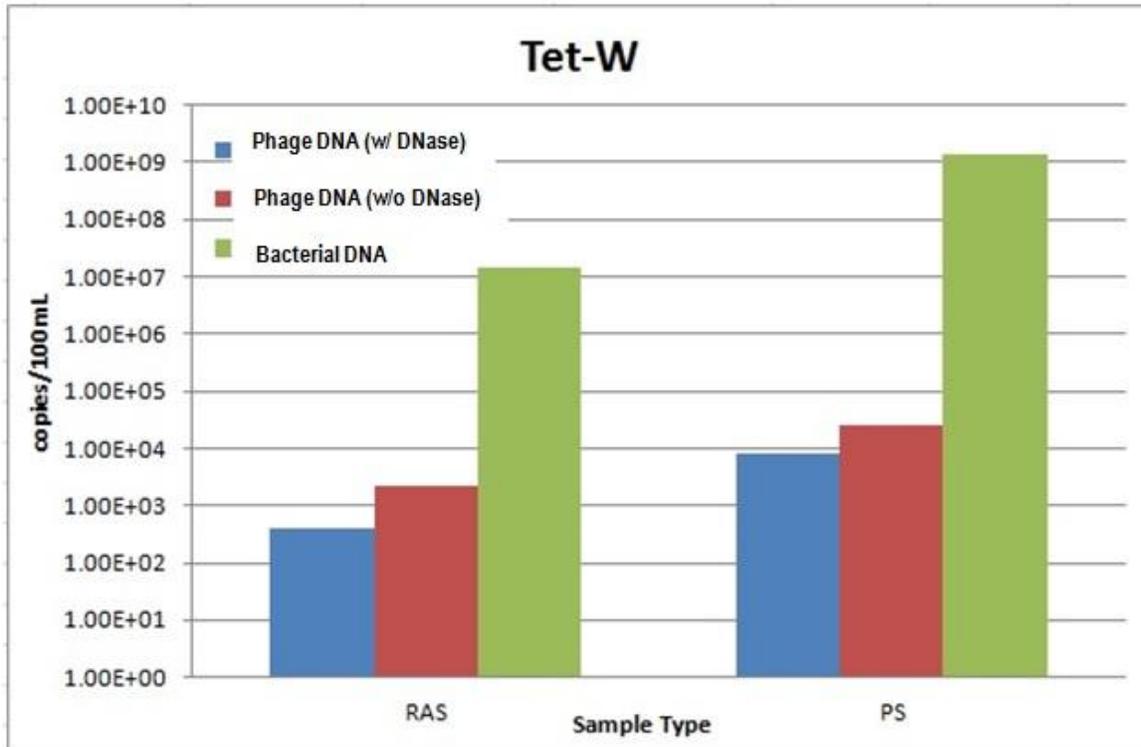


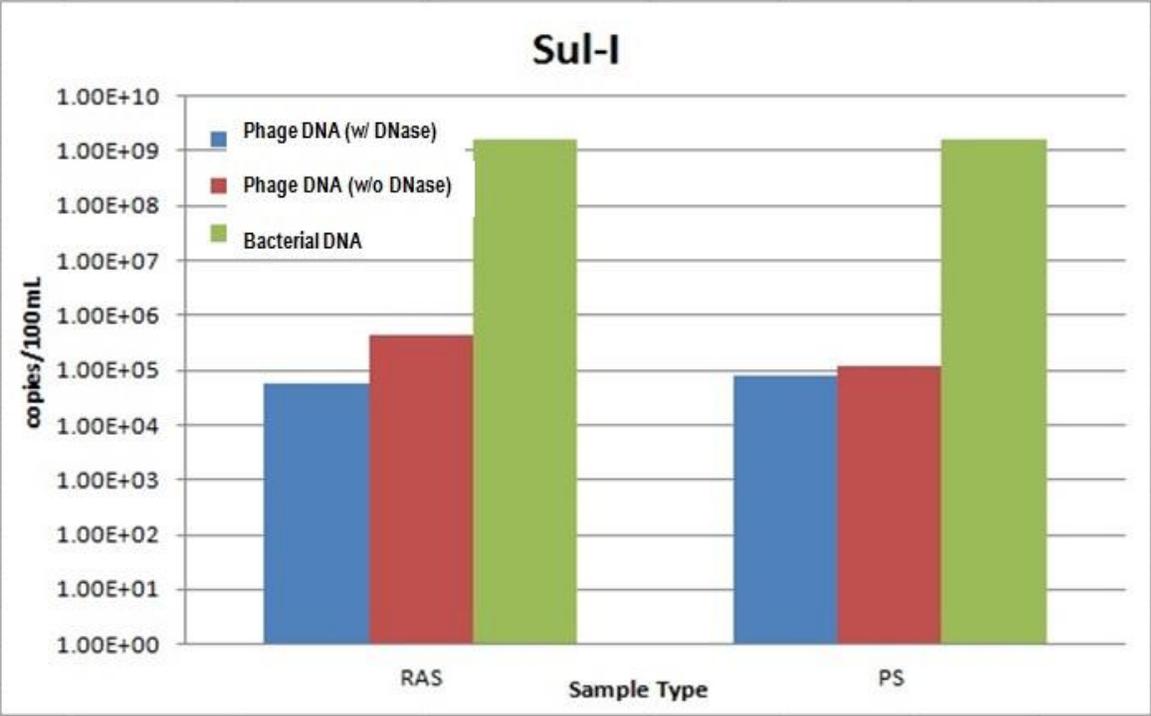
Figure 2.2: Organism (genus) Tree. The data was compared to M5NR using a maximum e -value of $1e-5$, a minimum identity of 60 %, and a minimum alignment length of 15 measured in *aa* for protein and *bp* for RNA databases. Color shading of the names indicates **genus** membership. Domain: viruses



(a)

Figure 2.3: Concentration (copies/100mL) of (a) tetracycline resistant gene (Tet W), and (b) sulfonamide resistant gene (Sul I) abundance in Phage DNA from sludge samples. Note: DNase indicate purified phage DNA after DNase treatment; Bacterial indicate overall bacterial DNA in the sample; RAS= Returned activated sludge, PS=Primary Sludge

Figure 2.3 (cont'd)



(b)

REFERENCES

REFERENCES

1. Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R., and Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *Journal of microbiological methods*, 92(2), 183-188.
2. Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J. and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology* 8(4), 251-259.
3. Aminov, R., Garrigues-Jeanjean, N. and Mackie, R. (2001). Molecular ecology of tetracycline resistance: development and validation of primers for detection of tetracycline resistance genes encoding ribosomal protection proteins. *Applied and environmental microbiology* 67(1), 22-32.
4. Auerbach, E.A., Seyfried, E.E. and McMahon, K.D. (2007). Tetracycline resistance genes in activated sludge wastewater treatment plants. *Water research* 41(5), 1143-1151.
5. Baquero, F., Martinez, J.L. and Canton, R. (2008) Antibiotics and antibiotic resistance in water environments. *Curr Opin Biotechnol* 19(3), 260-265.
6. Blasco, M.D., Esteve, C. and Alcaide, E. (2008). Multiresistant waterborne pathogens isolated from water reservoirs and cooling systems. *Journal of Applied Microbiology* 105(2), 469-475.
7. Borjesson, S., Melin, S., Matussek, A. and Lindgren, P.E. (2009). A seasonal study of the *mecA* gene and *Staphylococcus aureus* including methicillin-resistant *S. aureus* in a municipal wastewater treatment plant. *Water Res* 43(4), 925-932.
8. Boyd, E.F. and Brüssow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *TRENDS in Microbiology* 10(11), 521-529.
9. Brooks, J.B.J., Maxwell, S.M.S., Rensing, C.R.C., Gerba, C.G.C. and Pepper, I.P.I. (2007). Occurrence of antibiotic-resistant bacteria and endotoxin associated with the land application of biosolids. *Can J Microbiol* 53(5), 616-622.
10. Canchaya, C., Fournous, G. and Brüssow, H. (2004). The impact of prophages on bacterial chromosomes. *Molecular microbiology* 53(1), 9-18.
11. Clokie, M. R., Millard, A. D., Letarov, A. V., and & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, 1(1), 31-45.
12. Colomer-Lluch, M., Imamovic, L., Jofre, J. and Muniesa, M. (2011b). Bacteriophages carrying antibiotic resistance genes in fecal waste from cattle, pigs, and poultry. *Antimicrobial agents and chemotherapy* 55(10), 4908-4911.

13. Colomer-Lluch, M., Jofre, J. and Muniesa, M. (2011a). Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One* 6(3), e17549.
14. Guardabassi, L., Lo Fo Wong, D. and Dalsgaard, A. (2002). The effects of tertiary wastewater treatment on the prevalence of antimicrobial resistant bacteria. *Water research* 36(8), 1955-1964.
15. Kim, S. and Aga, D.S. (2007). Potential ecological and human health impacts of antibiotics and antibiotic-resistant bacteria from wastewater treatment plants. *Journal of Toxicology and Environmental Health, Part B* 10(8), 559-573.
16. Kim, S., Park, H. and Chandran, K. (2010). Propensity of activated sludge to amplify or attenuate tetracycline resistance genes and tetracycline resistant bacteria: A mathematical modeling approach. *Chemosphere* 78(9), 1071-1077.
17. Knapp, C. W., McCluskey, S. M., Singh, B. K., Campbell, C. D., Hudson, G., and Graham, D. W. (2011). Antibiotic resistance gene abundances correlate with metal and geochemical conditions in archived Scottish soils. *PLoS One*, 6(11), e27300.
18. Knapp, C.W., Dolfing, J., Ehlert, P.A. and Graham, D.W. (2009). Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environmental science & technology* 44(2), 580-587.
19. Luo, Y., Yang, F., Mathieu, J., Mao, D., Wang, Q., and Alvarez, P. J. J. (2013). Proliferation of Multidrug-Resistant New Delhi Metallo- β -lactamase Genes in Municipal Wastewater Treatment Plants in Northern China. *Environmental Science and Technology Letters*, 1(1), 26-30.
20. Mazaheri Nezhad Fard, R., Barton, M.D. and Heuzenroeder, M.W. (2011). Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Lett Appl Microbiol* 52(6), 559-564.
21. Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R. and Wilke, A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* 9(1), 386.
22. Muniesa, M., Colomer-Lluch, M., and Jofre, J. (2013a). Could bacteriophages transfer antibiotic resistance genes from environmental bacteria to human-body associated bacterial populations? *Mobile genetic elements*, 3(4).
23. Muniesa, M., Colomer-Lluch, M., and Jofre, J. (2013b). Potential impact of environmental bacteriophages in spreading antibiotic resistance genes. *Future microbiology*, 8(6), 739-751.
24. Muniesa, M., García, A., Miró, E., Mirelis, B., Prats, G., Jofre, J. and Navarro, F. (2004) Bacteriophages and diffusion of β -lactamase genes. *Emerging infectious diseases* 10(6), 1134.

25. Munir, M. and Xagorarakis, I. (2011). Levels of Antibiotic Resistance Genes in Manure, Biosolids, and Fertilized Soil. *Journal of Environment Quality* 40(1), 248.
26. Munir, M., Wong, K. and Xagorarakis, I. (2011). Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. *Water Res* 45(2), 681-693.
27. Nagulapally, S.R., Ahmad, A., Henry, A., Marchin, G.L., Zurek, L. and Bhandari, A. (2009). Occurrence of ciprofloxacin-, trimethoprim-sulfamethoxazole-, and vancomycin-resistant bacteria in a municipal wastewater treatment plant. *Water Environment Research* 81(1), 82-90.
28. Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper, W.F., Jr. and Liles, M.R. (2010). Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl Environ Microbiol* 76(11), 3753-3757.
29. Pei, R., Kim, S.-C., Carlson, K.H. and Pruden, A. (2006). Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water research* 40(12), 2427-2435.
30. Peltier, E., Vincent, J., Finn, C., and Graham, D. W. (2010). Zinc-induced antibiotic resistance in activated sludge bioreactors. *Water research*, 44(13), 3829-3836.
31. Prescott, J.F. (2004). Antimicrobial chemotherapy. In *Veterinary Microbiology* ed. Hirsh, D.C., Maclachlan, N.J. and Walker, R.L. Ames: Blackwell Publishing. 26–43.
32. Pruden, A., Pei, R., Storteboom, H. and Carlson, K.H. (2006). Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environmental science and technology* 40(23), 7445-7450.
33. Quirós, P., Colomer-Lluch, M., Martínez-Castillo, A., Miró, E., Argente, M., Jofre, J., Navarro, F., and Muniesa, M. (2014). Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Human Fecal Samples. *Antimicrobial agents and chemotherapy*, 58(1), 606-609.
34. Reinthaler, F., Posch, J., Feierl, G., Wüst, G., Haas, D., Ruckebauer, G., Mascher, F. and Marth, E. (2003) Antibiotic resistance of *E. coli* in sewage and sludge. *Water research* 37(8), 1685-1690.
35. Sander, M. and Schmieger, H. (2001). Method for host-independent detection of generalized transducing bacteriophages in natural habitats. *Appl Environ Microbiol* 67(4), 1490-1493.
36. Schwartz, T., Kohnen, W., Jansen, B., and Obst, U. (2003). Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water biofilms. *FEMS Microbiology Ecology*, 43(3), 325-335.

37. Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for de novo assembly of microbial genomes. *PLoS One*, 7(9), e42304.
38. Volkmann, H., Schwartz, T., Bischoff, P., Kirchen, S., and Obst, U. (2004). Detection of clinically relevant antibiotic-resistance genes in municipal wastewater using real-time PCR (TaqMan). *Journal of microbiological methods*, 56(2), 277-286.
39. Weinbauer, M.G. and Rassoulzadegan, F. (2003). Are viruses driving microbial diversification and diversity? *Environmental microbiology* 6(1), 1-11.
40. Wright, M. S., Peltier, G. L., Stepanauskas, R., and McArthur, J. V. (2006). Bacterial tolerances to metals and antibiotics in metal-contaminated and reference streams. *FEMS microbiology ecology*, 58(2), 293-302.
1. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, G. (1970). Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology*, 40(3), 734-744.
2. Zhang, X.-X. and Zhang, T. (2011). Occurrence, abundance, and diversity of tetracycline resistance genes in 15 sewage treatment plants across China and other global locations. *Environmental science and technology* 45(7), 2598-2604.
3. Zhang, Y., Marrs, C.F., Simon, C. and Xi, C. (2009). Wastewater treatment contributes to selective increase of antibiotic resistance among *Acinetobacter* spp. *Science of the Total Environment* 407(12), 3702-3706.

CHAPTER 3

METAGENOMIC INSIGHTS INTO MICROBIAL RESISTANCE TO ANTIBIOTIC AND METAL COMPOUNDS IN WASTEWATER UTILITIES

Mariya Munir, Terence Marsh, and Irene Xagorarakis. (in preparation)

Abstract

Over the past few years resistance to antibiotics has increased. Co-existence of antibiotics and metals may increase antibiotic resistance gene development in the environment. Wastewater treatment plants (WWTPs) can be considered as important reservoirs for the spread of antibiotic resistance to opportunistic pathogens and can stimulate horizontal gene transfer among microbial species. Metal exposure can directly select for metal-resistant bacteria while co-selecting for antibiotic-resistance. This study aimed to describe and compare the diversity and abundance of antibiotics and metal resistance in a conventional and MBR (membrane bioreactor) utility using metagenomic investigations. Illumina HiSeq sequencing was applied on six samples from two different WWTPs in Michigan. Bacterial DNA was isolated from three different sampling points (activated sludge (AS), before disinfection effluent (BD) and after disinfection effluent (AD)) from a conventional and MBR utility.

Sequencing reads from all the samples revealed differences in the abundance of functional genes within the WWTPs. Genes coding for antibiotic resistance were identified in all the samples. Most of the antibiotic resistance genes conferred resistance to fluoroquinolones, beta-lactamase, methicillin, and erythromycin and vancomycin. Genes coding for resistance to metals were also observed in all our samples. High resistance to metals (including Cobalt-zinc-cadmium resistance, zinc resistance, arsenic resistance, copper tolerance, and resistance to

chromium compounds, mercury and cadmium resistance) was detected in most of the samples. The MBR utility showed slightly higher number of hits for all the functional categories compared to the conventional WWTP samples. The incidence of multiple metal and antibiotic resistances among bacterial populations in WWTP poses a potential threat to human health.

Keywords: Metagenomics, wastewater, effluents, activated sludge, antibiotic resistance, metal resistance, Illumina Hiseq

1. Introduction

The occurrence of antibiotic resistance bacteria (ARB) and antibiotic resistant genes (ARG) in our environment is a growing global health problem. The microbial quality of water itself is of great concern; however, if the trace levels of antibiotics and antibiotic resistant bacteria are present, they may also greatly affect public health and is an emerging issue for the general public and water industries (Xi et al. 2009). Development of novel antibiotics is being outpaced by rapid propagation of antibiotic resistance thus calling for effective strategies to mitigate the spread of antibiotic resistance (Carlet et al., 2012).

Wastewater treatment plants (WWTPs) can be considered as an important reservoirs for the spread of antibiotic resistance to opportunistic pathogens and can stimulate horizontal gene transfer among microbial species. Large amounts of antibiotics are released into municipal wastewater due to incomplete metabolism in humans or to disposal of unused antibiotics. Once in the wastewater stream they can exert selective pressure for or maintain resistance among microorganisms (Allen et al. 2010, Nagulapally et al. 2009).

Due to the increasing evidence of antibiotic resistance in pathogenic and benign bacteria in our environment, an emerging threat to public and environmental health has been postulated (LaPara et al., 2011; Munir and Xagorarakis 2011, Knapp et al. 2009, Blasco et al. 2008).

Antibiotic resistance genes (ARGs) which code for specific antimicrobial functions such as efflux pumps is considered to play a major role in conferring antibiotic resistance to microbial community (Webber and Piddock, 2003). Antibiotic resistant bacteria and genes encoding antibiotic resistance are commonly detected at high rates and concentrations in wastewater samples (Munir et al. 2011, Zhang and Zhang 2011, Borjesson et al. 2009, Zhang et al. 2009, Auerbach et al. 2007, Brooks et al. 2007, Kim and Aga 2007, Pruden et al. 2006). A strong link has been reported between wastewater and antibiotic resistance (Börjesson S et al. 2009, Volkmann et al. 2004, Reinthaler et al. 2003, Guardabassi et al. 2002).

Disinfection methods including chlorine or UV disinfection are capable of reacting with nucleic acids during treatment and therefore may potentially reduce ARGs (Dodd, 2012).

Previous studies have shown that disinfection process did not contribute much in the reduction of ARGs and ARBs in wastewater effluents from full-scale WWTPs (Fahrenfeld et al. 2013; Munir and Xagorarakis 2011, Auerbach et al., 2007). Studies have suggested that multidrug resistant genes even survive through several wastewater treatment units, including disinfection (Luo et al. 2014, Shi et al. 2013, Odjadjare et al. 2012). In a recent controlled lab study led by Mckinney and Pruden (2013), it was demonstrated that UV disinfection at WWTPs is capable of reducing strains of bacteria that are resistant to antibiotics but not ARGs.

Heavy metals along with antibiotics also create a selective pressure in the environment that leads to resistance (Baquero et al. 1998). It is believed that multiple stresses can provide more ecologically favorable conditions for a bacterium to survive and acquire resistance in an environment, for example antibiotics and heavy metals (Spain and Alm 2003). Recently it has been reported that chemical contaminants like metals can influence the selection of antibiotic resistance among bacteria (Deredjian et al. 2011, Chandra and Sankhwar 2011). Sub-toxic levels

of zinc have shown to increase antibiotic resistance in wastewater treatment plant's microbial communities at comparatively low levels of antibiotic. The reason for such an observation was suggested to be development of cross-resistance (Peltier et al. 2010). Presence of metals in wastewater treatment can also be one of the factors responsible for selection of antibiotic resistance among exposed bacteria in the environment (Peltier et al. 2010, Knapp et al. 2011, Kamala-Kannan and Lee 2008, Tuckfield and McArthur 2008, Baker-Austin et al. 2006, Wright et al. 2006, Stepanauskas et al. 2005). Wright et al. (2006) concluded that metal exposure can directly select for metal-resistant bacteria while co-selecting for antibiotic-resistant bacteria. Recently one study has shown that ARG concentration significantly correlate with presence of metals (Knapp et al. 2011).

Next-generation sequencing (NGS) has substantially widened the scope of metagenomic analysis of environmentally derived samples (Mardis 2008). Studies have demonstrated NGS to be effective in detecting gene modifications responsible for antibiotic resistance (La Scola et al. 2008). High-throughput sequencing was used successfully to highlight the prevalence of ARGs and mobile genetic elements in microbial population of sewage treatment plants and is considered as a promising tool for analyzing ARG and other functional diversity in the environmental samples (Zhang et al. 2011).

The objective of this study was to describe and compare the diversity of microbial resistances to antibiotics and metal compounds in a conventional and MBR (membrane bioreactor) utility using metagenomic investigations. This study provided a bioinformatics approach for identifying microbial population and functional features like antibiotic resistance and metal resistance in wastewater utilities. This is the first study investigating microbial resistance patterns along with metal resistance in bacterial isolated samples in WWTPs.

2. Materials and Methods

2.1. Sample Collection: Effluent samples before and after disinfection along with activated sludge were collected from East Lansing (EL) WWTP and Traverse City (TC) WWTP in Michigan (U.S.A.) in 2013. Table 3.1 provides the characteristics of these WWTPs based on wastewater treatment processes and disinfection methods. The schematic of the sampling along with location of WWTPs is shown in Figure 3.1. The schematic of the methods used in this study is also presented in Figure 3.2. Two liters of grab effluent samples were collected for bacterial isolation from each WWTP. Activated sludge (AS) sample was collected in two 1L nalgene bottles, mixed together in laboratory and then bacteria isolation from each WWTP was done. Samples were kept on ice and were transported to the Water Quality Engineering Laboratory at Michigan State University (East Lansing, U.S.A.) for further immediate processing.

2.2. Sample Processing: Bacteria in the effluent samples were concentrated by filtration with 0.45 μm HA filters (Millipore, Billerica, MA). The volume of effluent samples filtered was 1 L for each of the four samples. The filters were collected in a 50 mL tubes and 50 mL Phosphate Buffer Water (PBW) was added in each tube containing a filter. The tubes were then vortexed for 5 min to allow the biomass layer on the filters to mix with water. 50mL of AS samples were also collected in a centrifuge tubes. All the tubes were then centrifuged for 20 min at 4500 rpm to concentrate the sample down to 2 mL. Supernatant was discarded and the concentrates were stored at $-80\text{ }^{\circ}\text{C}$ until the DNA extraction was performed for further molecular analysis.

2.3. Nucleic acid Extraction: Bacterial DNA was extracted (from samples before and after disinfection) using a MagNA Pure Compact DNA extractor (Roche Applied Science, Indianapolis, IN, USA) following the protocol in the manufacturer's manual. The MagNA Pure Compact utilizes a magnetic-bead technology for the isolation process. Sample amount of 400

μL was loaded in the system and the elution volume was $100\mu\text{L}$. The extracts were stored in a freezer at -20°C . Following extraction the quantity of bacterial nucleic acid extracts from all samples were checked using the NanoDrop Spectrophotometer (NanoDrop[®] ND-1000, Wilmington, DE).

2.5. Metagenomic sequencing and analyses: All samples including Bacterial enriched DNA isolated from the sludge and the effluent samples was sequenced on an Illumina platform (Illumina HiSeq, Roche Technologies) at The Research Technology Support Facility (RTSF) genomic center at Michigan State University generating 250 bp paired-end reads. Approximately 1 μg DNA (per core sample) was sent to the sequencing facility. The sequencing results were returned as .fastq.gz files. Further files were converted to fastq files by processing them in MSU HPCC (High Performance Computing Center) secure shell (SSH) connection. PuTTY, a freely available piece of software was used to establish this SSH connection. In order to quality filter the illumina data, a flexible read trimming tool for Illumina NGS data called Trimmomatic was used for trimming Illumina data and removing adapters (Bolget et al. 2014). Finally, the trimmed sequences were assembled using an iterative De Bruijn Graph De Novo Assembler for Short Reads Sequencing data with Highly Uneven Sequencing Depth called **IDBA-UD** (Peng et al. 2012).

2.6. MG-RAST analyses: The assembled data from all the samples were analyzed using MetaGenome Rapid Annotation Subsystems Technology server (MG-RAST 3.3.1) (Meyer et al. 2008). Each assembled data file underwent quality control (QC) process, which included quality filtering (removing sequences with ≥ 5 ambiguous base pairs), length filtering (removing sequences with a length ≥ 2 standard deviations from the mean), and de-replication (removing similar sequences that are artifacts of sequencing). Analysis includes the phylogenetic

comparisons and functional annotations against the database and the results are expressed in the form of abundance profiles. The “abundance” in the MG-RAST presents the estimate of the number of sequences that contain a given annotation, found by multiplying each selected database match (hit) by the number of representatives in each cluster. Hits refer to the number of unique database sequences that were found in the similarity search, not the number of reads. The hit count can be smaller than the number of reads because of clustering or larger due to double counting (Wilke et al. 2014). Data was analyzed based on organism abundance and on the functional distribution at the subsystem hierarchy with maximum E- value cutoff of $1E^{-5}$, minimum percent identity cutoff of 60% and minimum alignment length cutoff of 15bps. The MG-RAST pipeline analysis includes the comparisons and functional annotations against the database. Each of the categories was further studied for detailed analysis and data was downloaded in excel sheet for further analysis.

2.7. Statistical analysis: Abundance data was downloaded from MGRAST and statistical analysis was performed using Microsoft Office Excel. Relative abundance is defined as the number of sequences mapped to specific function divided by the total number of sequences in that sample. The relative abundance data is then normalized using the following formula:

The normalized value of e_i for variable E in the i^{th} row is calculated as:

$$\text{Normalized } (e_i) = \frac{e_i - E_{\min}}{E_{\max} - E_{\min}} \quad \text{Where } E_{\min} = \text{the minimum value for variable E and}$$

E_{\max} = the maximum value for variable E.

3. Results and Discussion:

3.1. Metagenome sequencing and assembly results:

To analyze the wastewater metagenomes, Illumina Hiseq sequencing was applied on samples from two different WWTPs in Michigan. Bacterial DNA was isolated from three different sampling points (activated sludge (AS), before disinfection effluent (BD) and after disinfection effluent (AD)) from a conventional and MBR utility. Sequences generated were assembled using an IDBA-UD assembler. Metagenome analysis revealed that a total of 2355 Mbp (mega base pairs) of assembled sequence data was generated. Bacterial DNA samples contained contigs ranging from 238657 to 380106 sequences totaling 1250 Mbp of sequences.

Exploring the metagenome sequence breakdown, it was found that 6,741 sequences (1.9%) failed to pass the QC pipeline for EL activated sludge sample. Of the sequences that passed QC, 460 sequences (0.1%) contain ribosomal RNA genes. Of the remainder, 223,213 sequences (64.0%) contain predicted proteins with known functions and 117,626 sequences (33.7%) contain predicted proteins with unknown function. 795 (0.2%) of the sequences that passed QC have no rRNA genes or predicted proteins. Out of the 342,096 sequences (totaling 204,012,220 bps) that passed quality control, 340,839 (99.6%) produced a total of 391,844 predicted protein coding regions. Of these 391,844 predicted protein features, 213,941 (54.6% of features) have been assigned an annotation using at least one of our protein databases (M5NR) and 165,694 features (77.4% of annotated features) were assigned to functional categories. For TC sludge sample, 7,927 sequences (2.1%) failed to pass the QC pipeline. Of the sequences that passed QC, 543 sequences (0.1%) contain ribosomal RNA genes. Of the remainder, 276,740 sequences (72.8%) contain predicted proteins with known functions and 94,297 sequences

(24.8%) contain predicted proteins with unknown function. 597 (0.2%) of the sequences that passed QC have no rRNA genes or predicted proteins. Of the 372,179 sequences (totaling 227,949,491 bps) that passed quality control, 371,037 (99.7%) produced a total of 441,132 predicted protein coding regions. Of these 441,132 predicted protein features, 266,439 (60.4% of features) have been assigned an annotation using at least one of our protein databases (M5NR) and 215,875 features (81.0% of annotated features) were assigned to functional categories.

In before disinfection effluent samples from EL, 5,765 sequences (2.2%) failed to pass the QC pipeline. Of the sequences that passed QC, 702 sequences (0.3%) contain ribosomal RNA genes. Of the remainder, 176,685 sequences (68.8%) contain predicted proteins with known functions and 70,878 sequences (27.6%) contain predicted proteins with unknown function. 2,845 (1.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Within 251,116 sequences (totaling 153,515,277 bps) that passed quality control, 247,563 (98.6%) produced a total of 297,712 predicted protein coding regions. Of these 297,712 predicted protein features, 167,031 (56.1% of features) have been assigned an annotation using at least one of the protein databases (M5NR) and 137,654 features (82.4% of annotated features) were assigned to functional categories. Similarly, in bacterial isolated effluent sample after disinfection from EL, 4,302 sequences (1.8%) failed to pass the Quality Control (QC) pipeline. Of the sequences that passed QC, 841 sequences (0.4%) contain ribosomal RNA genes as computed by MGRAST analysis software. Of the remainder, 184,215 sequences (77.2%) contain predicted proteins with known functions and 48,954 sequences (20.5%) contain predicted proteins with unknown function. 339 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Out the 234,355 sequences (totaling 147,771,487 bps) that passed quality control, 233,169 (99.5%) produced a total of 270,966 predicted protein coding regions.

Of these 270,966 predicted protein features, 172,766 (63.8% of features) have been assigned an annotation using at least one of the protein databases (M5NR) and 148,347 features (85.9% of annotated features) were assigned to functional categories.

In TC, before disinfection effluent samples, 6,703 sequences (2.6%) failed to pass the QC pipeline. Of the sequences that passed QC, 552 sequences (0.2%) contain ribosomal RNA genes. Of the remainder, 202,533 sequences (78.1%) contain predicted proteins with known functions and 49,351 sequences (19.0%) contain predicted proteins with unknown function. 77 (0.0%) of the sequences that passed QC have no rRNA genes or predicted proteins. Whereas in effluent sample after disinfection from TC, 3,695 sequences (1.3%) failed to pass the QC pipeline. Of the sequences that passed QC, 624 sequences (0.2%) contain ribosomal RNA genes. Of the remainder, 231,252 sequences (80.0%) contain predicted proteins with known functions and 53,202 sequences (18.4%) contain predicted proteins with unknown function. 153 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Detailed metagenome statistics for all the samples is shown in Table 3.2 a & b.

3.2. Resistance to antibiotics: Antibiotic Resistant Gene Diversity:

All six wastewater metagenomes were searched for functional signatures of resistance genes using MGRAST. Around three percent (3.6-3.8%) of the sequences of the bacterial enriched metagenomic library could be mapped to “virulence, disease and defense” genes using the subsystem functional database classification in MGRAST server. The ‘virulence, disease and defense’ functional category was analyzed further for resistance to antibiotic and toxic compounds category. In all the datasets of bacterial samples from TCWWTP and ELWWTPs, 74-76% of sequences were associated with resistance to antibiotic and toxic compounds

(RATC). A schematic of the process followed to analyze these metagenomes is presented in Figure 3.3.

Going deeper into these categories for all the metagenomes, cobalt-zinc-cadmium resistance is the most frequently occurring functional group followed by multidrug resistance efflux pumps and resistance to fluoroquinolones. It has been reported that the fluoroquinolone resistance genes are of great concern as they may contribute to the development of resistance to this class of antibiotics in humans because of their use in veterinary medicine, particularly for food animals (Durso et al., 2011).

To further analyze these metagenomes, different resistance hits were grouped according to their function (Table 3.3) provided in supplement data. Analysis based on each grouping category was conducted and results are presented in Figure 3.4 to 3.6. Around 18-23% of RATC sequences in all bacterial samples were associated with resistance to antibiotics. Figure 3.4 shows the abundance of bacterial resistance to antibiotics in sludge samples from EL and TC WWTPs. It was found that most of the antibiotic resistance genes conferred resistance to fluoroquinolones, beta-lactamase, methicillin, and erythromycin and vancomycin group of antibiotics (Table 3.3). The level of abundance of resistance to antibiotics was higher in TC sludge samples compared to EL WWTP. Similar results were observed for all the effluent samples (Figure 3.5 and 3.6).

Also, the chlorination effects on microbial antibiotic resistance in a wastewater treatment plants were investigated. Using metagenomic analysis, no significant difference was observed in functional categories in before and after disinfection samples from both the treatment plant. In order to have a comparison between different samples, normalization of the sequence hits was

done by total number of sequences in each dataset based on functional grouping (Table 3.3). Figure 3.7 shows the relative abundance of different functional resistance in bacterial isolated samples. According to the relative abundance of each bacterial sample analyzed with MGRAST, no significant difference was observed between before and after disinfection samples from both ELWWTP and TCWWTP. Although a slight reduction in resistance is observed within TC WWTP. From the figure, it looks like disinfection further concentrated the antibiotic resistance and resistance to metal signatures. Similar higher proportion of resistance was observed after chlorination in drinking water treatment (Shi et al. 2013).

A significant percentage (18-24%) of metagenome sequences are associated with “multidrug resistance efflux pump” and were prevalent in all the bacterial samples (Figure 3.4 to 3.6). Efflux pumps play an important defensive role against different toxic compounds that bacteria may encounter in their environment. The role of efflux pumps antibiotic resistance is their ability to export antibiotics and other drugs out of bacterial cells (Fernandez and Hancock, 2012). Further, high-throughput sequencing was used successfully to highlight the prevalence of ARGs in wastewater samples. Metagenomic analysis can also help in identifying the novel ARG determinants within WWTPs.

3.3. Resistance to metal compounds:

Functional classifications of the metagenomic sequences for analyzing the resistance to metals were determined by annotating sequences to the functional gene database using MGRAST program. Different functional hits were grouped according to their function role and the grouping as shown in table (Table 3.3) provided in supplement data. Based on literature review, metals are usually categorized as either essential metals (important for life) or non essential

metals (with no known physiological functions to humans) (Krasnici et al., 2013; Ishaque et al. 2006; Francisco et al. 2002). Metals that are generally regarded as essential for human health in trace amounts include iron, zinc, copper, manganese. They are essential because they form an integral part of one or more enzymes involved in a metabolic or biochemical process. Non-essential metals (lead, mercury, cadmium) are of main concern pollutants because they pose potential risks to human health and the environment (Ishaque et al. 2006).

Further exploring the RATC gene category within all the metagenomes, cobalt-zinc-cadmium resistance is the most frequently occurring functional group in bacterial metagenomes. In all the datasets, 47-50% of these sequences were associated with resistance to metals. Figure 3.4, 3.5 and 3.6 shows the abundance of bacterial resistance to metals in sludge, before disinfection effluents and after disinfection effluent samples from EL and TC WWTPs respectively. An interesting high resistance to essential metals (including Cobalt-zinc-cadmium resistance, copper tolerance, and zinc resistance) was detected in all the samples. Around 5-8% of these samples were annotated to resistance to non-essential metals (resistance to arsenic, chromium compounds, mercury and cadmium). The level of abundance of Cobalt-zinc-cadmium resistance was usually higher in TC compared to ELWWTP for most of the samples.

Wright et al. (2006) concluded that metal exposure can directly select for metal-resistance while co-selecting for antibiotic-resistant bacteria. Contaminations due to heavy metals are widespread. As it has been suggested that presence of metals in wastewater treatment can also drive for selection of antibiotic resistance among bacteria, then they could be a more long-lasting source of resistance than are antibiotics themselves (Peltier et al. 2010; Knapp et al. 2011; Wright et al. 2006). Heavy metals cannot be degraded and thus, difficult to remove from the environment (Sinha et al. 2013). Presence of essential metals in high concentrations can be

toxic or can exert a selective pressure in bacterial communities. This selective pressure could also select for antibiotic resistant bacteria, and thus may play an important role in the maintenance and proliferation of antibiotic resistance (Francisco et al. 2013). According to Knapp et al. (2009), defense-associated metal resistance genes are often closely related with antibiotic resistance genes which can either encode for generic detoxifying mechanisms (e.g., efflux pumps) where intracellular concentrations of both metals and antibiotics is reduced non-specifically (cross resistance), or they may involve separate genes integrated on the same genetic element (co-resistance).

These results present a unique opportunity to examine the WWTPS with respect to diversity and presence of ARG along with other metal resistances, thus adding the knowledge gap to the number of metagenomic datasets previously available. Metagenome RATC data on different samples are important and can be further used to connect antibiotic and metal resistance information with microbial community in wastewater environment. In order to comprehensively and deeply characterize these metagenomes it is necessary to increase the sequencing depth. Although the lack of replication makes it difficult to draw wide conclusions regarding the effects treatment type have on microbial community about ARGs or metal resistances, these metagenomes provide important data to make baseline observations that will need to be examined more thoroughly in future studies.

4. Conclusions:

In the present work, we analyzed activated sludge and effluent samples using metagenomic analysis in an effort to better understand the composition and diversity of antibiotic resistance and resistance to metal compounds in wastewater utilities. Most of the antibiotic resistance genes conferred resistance to fluoroquinolones, beta-lactamase, methicillin, and

erythromycin and vancomycin group of antibiotics. The abundance of resistance to fluroquinones was higher in TC WWTP samples compared to EL WWTP.

High resistance to metals (including Cobalt-zinc-cadmium resistance, copper tolerance, zinc resistance, Arsenic, chromium compounds, mercury and cadmium resistance) was also detected in all samples. For sludge samples, MBR utility showed slightly higher number of hits for all the functional categories as described in the grouping. Further research is needed to provide better understanding of the comparison between treatments.

5. Acknowledgement:

We would like to thank the managers of the East Lansing Wastewater Treatment Plant and Traverse City Wastewater Treatment Plant for providing the samples and information needed for this study. We would like to acknowledge bioinformatic support and assistance provided by Bioinformatic team at High Performance Computing Center (HPCC) at Michigan State University.

APPENDIX

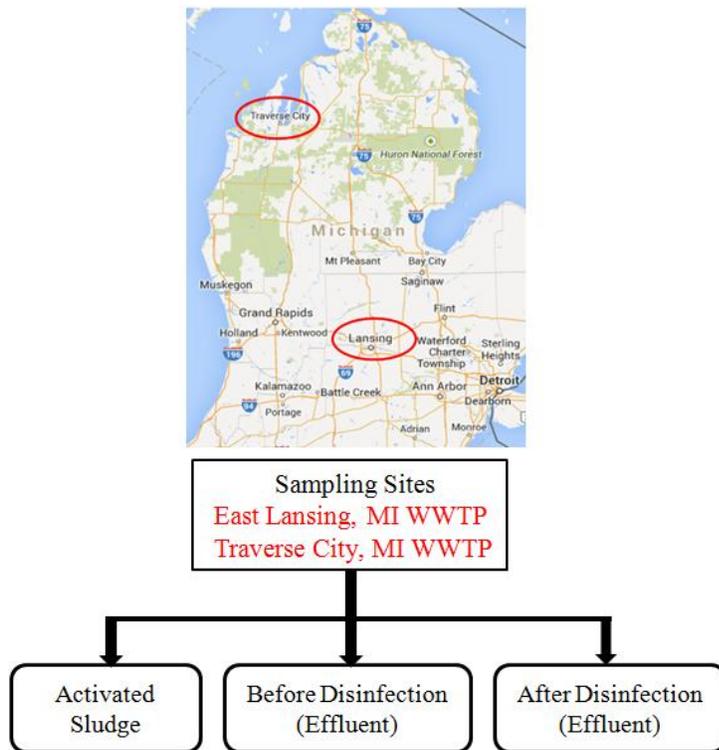


Figure 3.1: Map showing location of sampling and sampling schematic.

Table 3.1: Wastewater Treatment Characteristics

	EAST LANSING WWTP	TRAVERSE CITY WWTP
Wastewater treatment process (Biological treatment)	Activated Sludge (AS)	Membrane BioReactor (MBR)
Sludge Retention Time (SRT)	14 days	7.58 days
Capacity	18.8 MGD*	17.0 MGD
Average flow	13.4 MGD	8.5 MGD
Discharge Rate	14.1 MGD	4.0 MGD
Disinfection	<i>Chlorine (Cl)</i>	<i>Ultra-Violet (UV)</i>

*MGD-Millions gallon per day

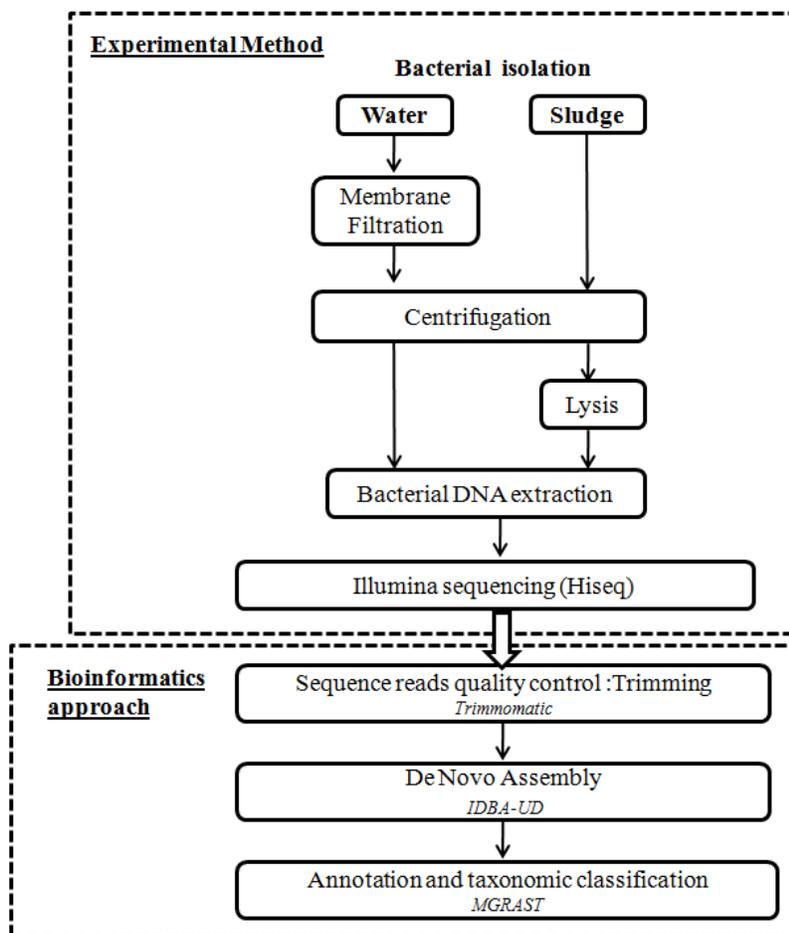


Figure 3.2: Schematic flowchart showing the procedure and methodology.

Table 3.2a: Metagenome analysis statistics in bacterial samples.

Metagenome Name	Raw bp count	No. of contigs	Mean Sequence Length (bp)	Post QC: bp Count	Post QC No. of contigs	Post QC: Mean Sequence Length (bp)
East Lansing Waste Water Treatment Plant						
EL_AS	237008412	348837	679 ± 1053	204012220	342096	596 ± 345
EL_BD	184817243	256881	719 ± 1374	153515277	251116	611 ± 411
EL_AD	176512330	238657	739 ± 1927	147771487	234355	630 ± 482
TC						
TC_AS	265323338	380106	698 ± 993	227949491	372179	612 ± 357
TC_BD	179323360	259220	691 ± 1048	14,645886	252517	584 ± 339
TC_AD	206030297	288930	713 ± 2198	176537661	285235	618 ± 484

Note: Sequences were assembled and contigs generated were analyzed on MGRAST; Abbreviation: bp= base pair

Table 3.2b: Functional category Hit distribution.

Metagenome Name	Processed: Predicted Protein Features	Processed: Predicted rRNA Features	Alignment :Identified Protein Features	Alignment :Identified rRNA Features	Annotation: Identified Functional Categories
East Lansing Waste Water Treatment Plant					
EL_AS	391,844	29,185	213,941	414	165,694
EL_BD	297,712	22,251	167,031	625	137,654
EL_AD	270,966	17,827	172,766	758	148,347
Traverse City Waste Water Treatment Plant					
TC_AS	441,132	28,367	266,439	500	215,875
TC_BD	297,052	19,752	191,448	499	156,342
TC_AD	341,222	23,529	218,046	546	179,026

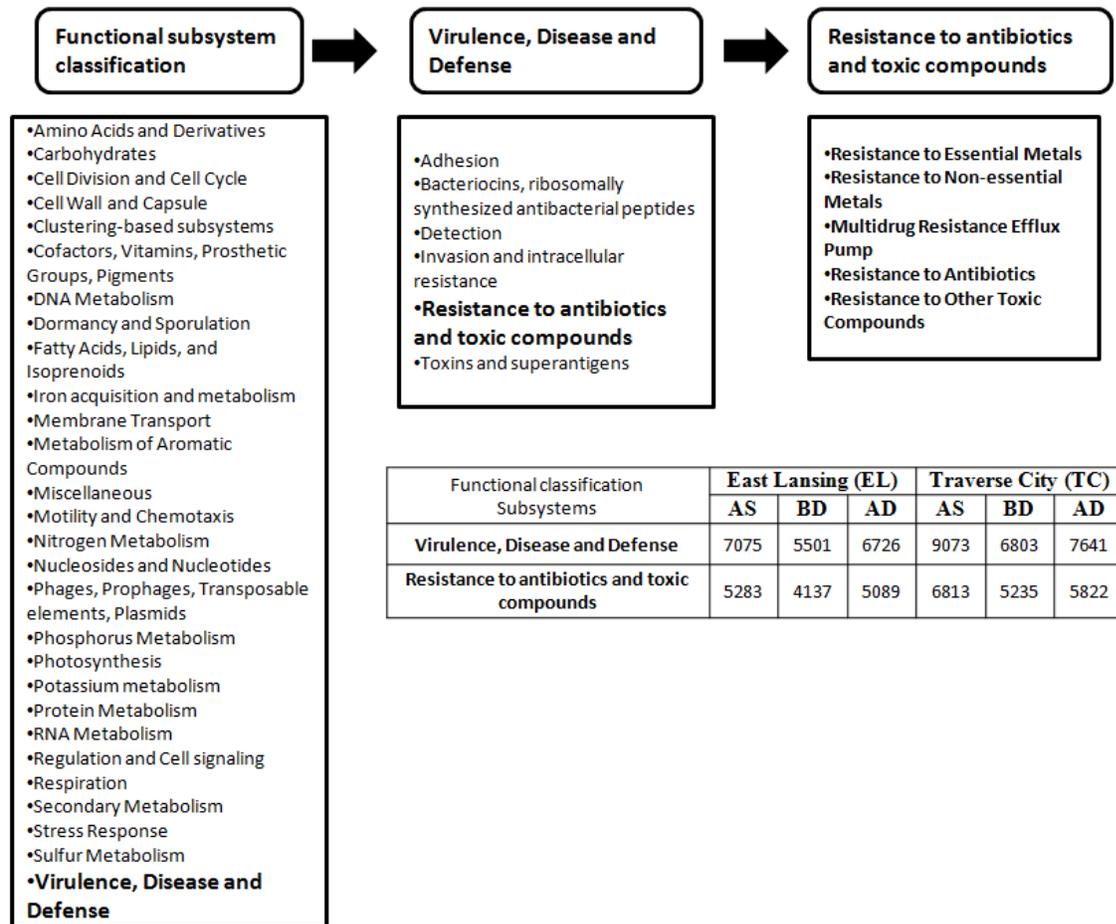


Figure 3.3: Schematic showing the analysis step for the functional abundance in MGRAST.

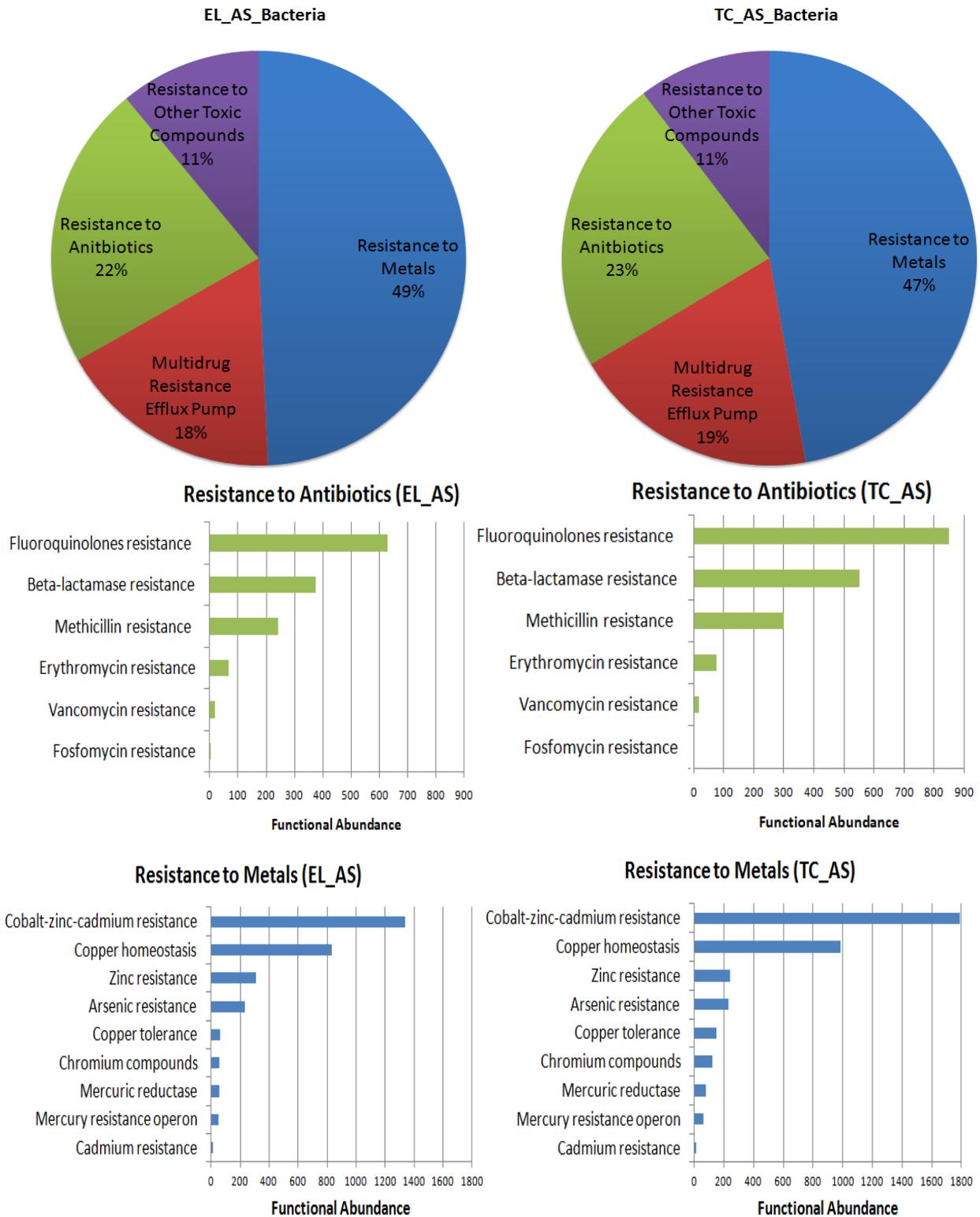


Figure 3.4: Bacterial resistance to antibiotics and metals in activated sludge for East Lansing and Traverse City wastewater utilities.

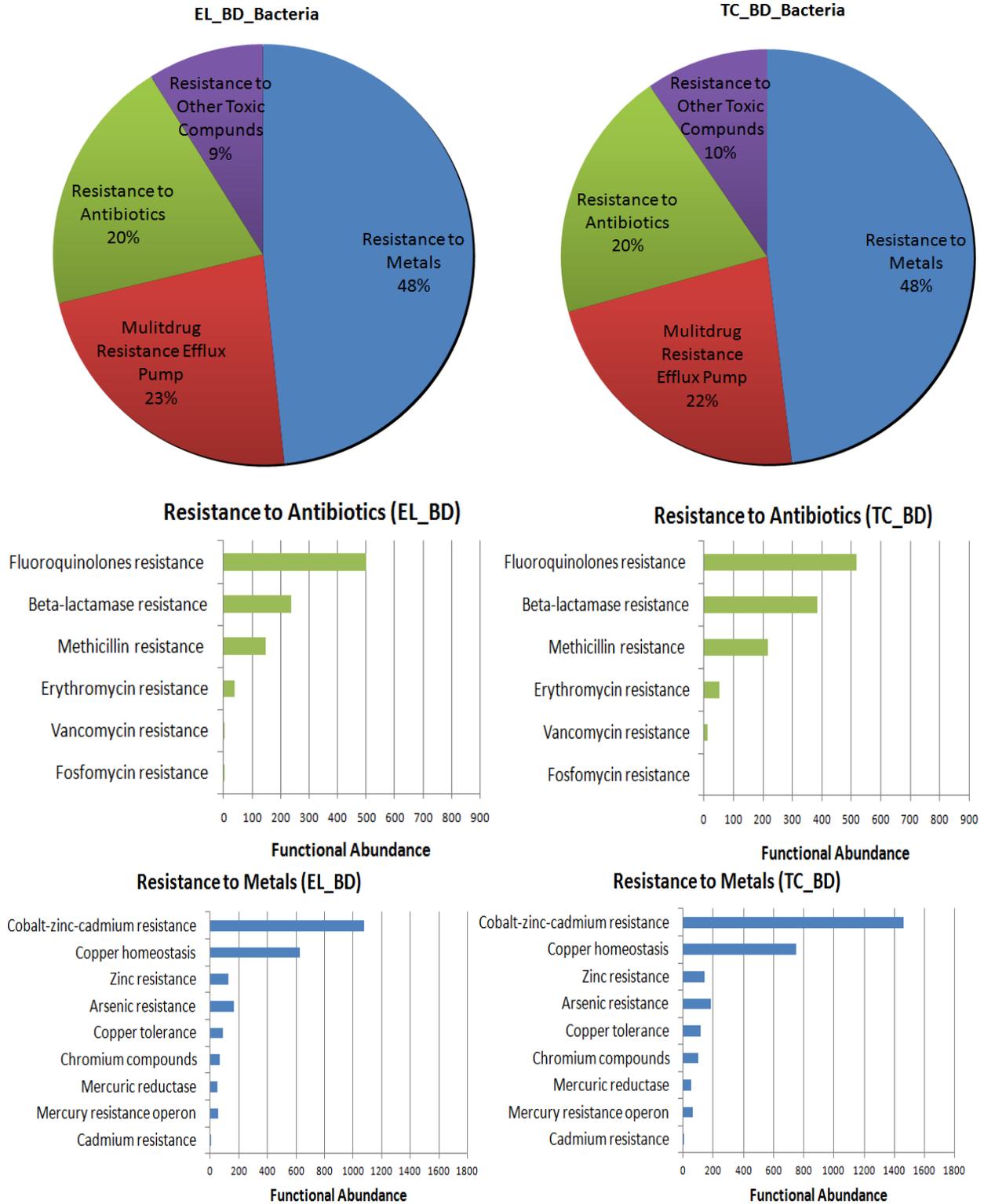


Figure 3.5: Bacterial resistance to antibiotics and metals in before disinfection effluent samples for East Lansing and Traverse City wastewater utility.

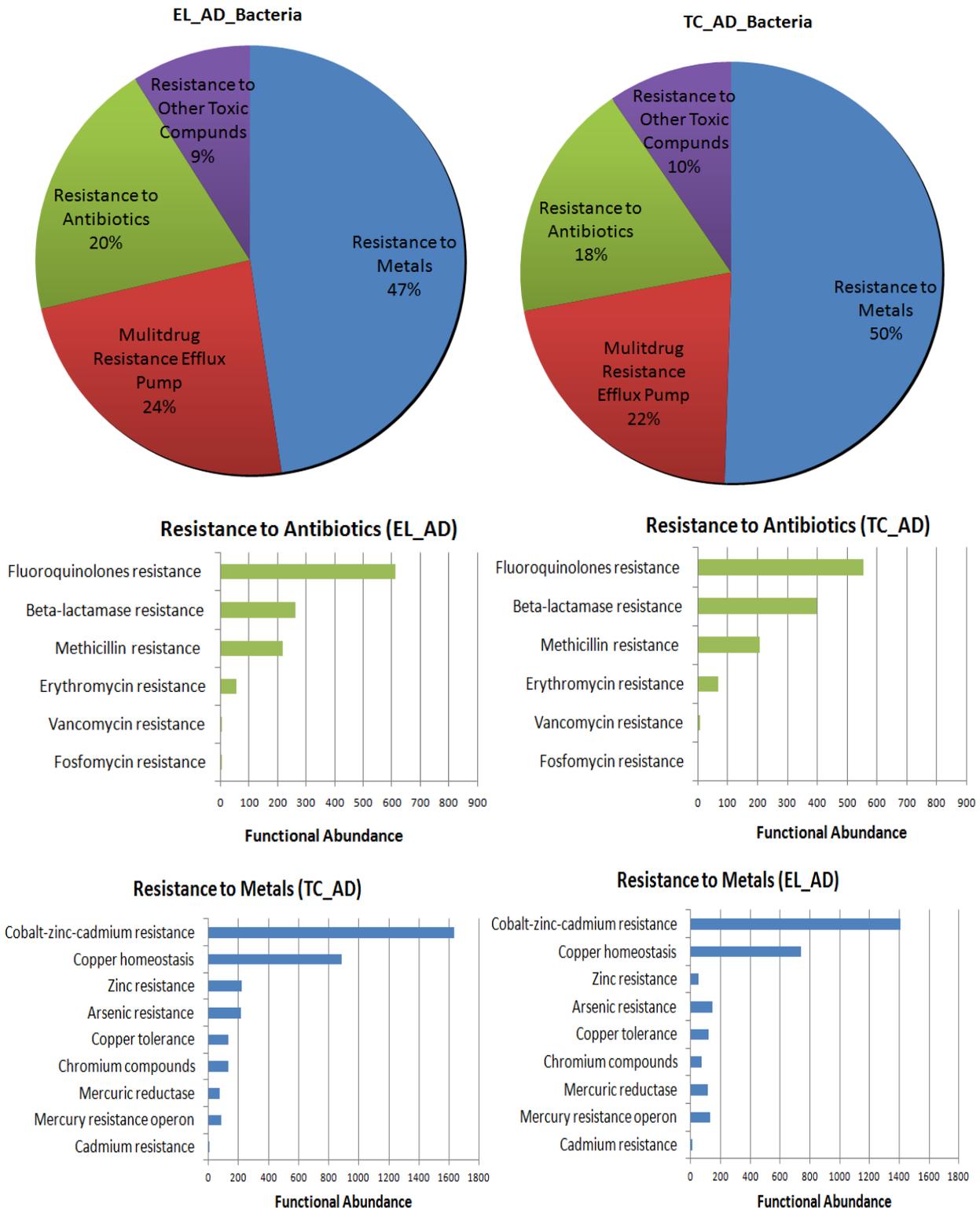


Figure 3.6: Bacterial resistance to antibiotics and metals in after disinfection effluent samples for East Lansing and Traverse City wastewater utility.

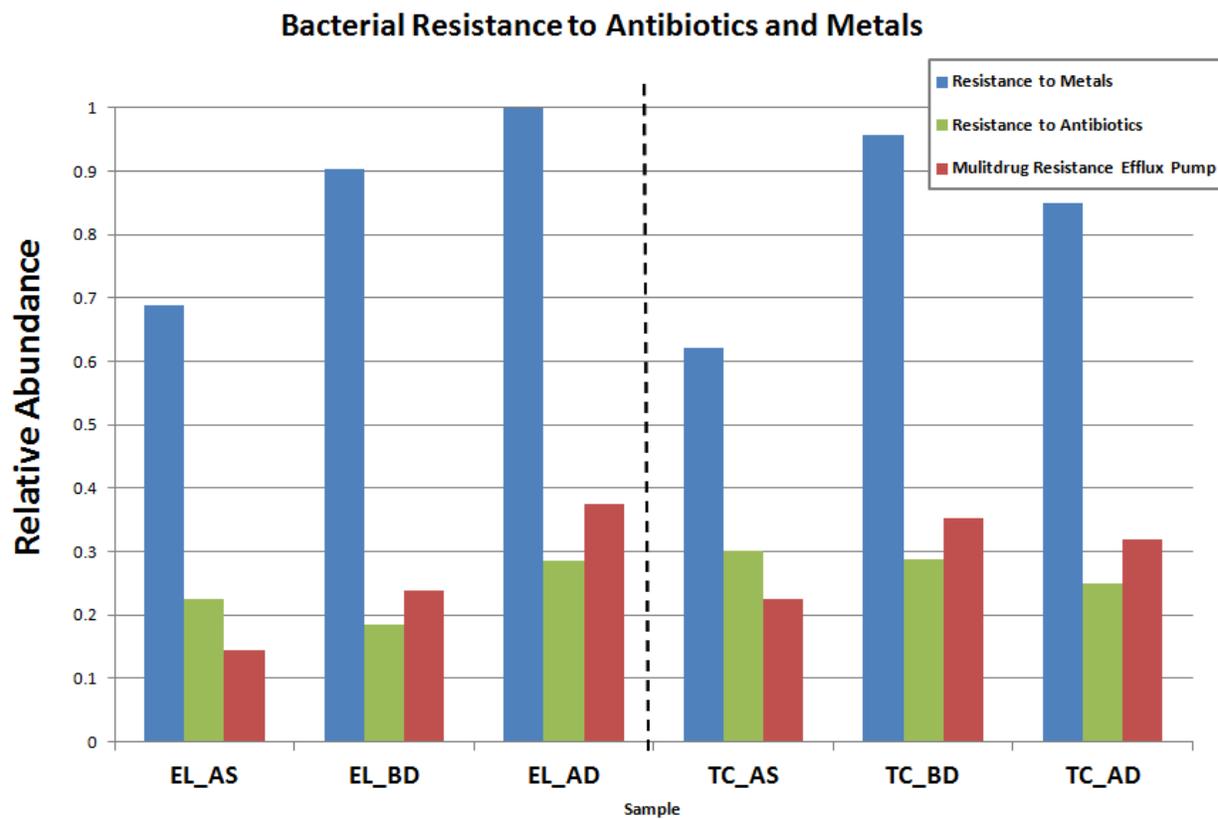


Figure 3.7: Relative abundance of resistance to antibiotics and metal compounds obtained from MGRAST.

Table 3.3: Abundance of Resistance groups for bacterial samples (MGRAST)

East Lansing and Traverse City – Before and After Disinfection and Activated Sludge (Bacteria)							
Resistance Groups	Antibiotic and Toxic Compound Resistance Distribution	East Lansing (EL)			Traverse City (TC)		
		AS	BD	AD	AS	BD	AD
Resistance to Metals	Cobalt-zinc-cadmium resistance	1400	1079	1406	1847	1460	1632
	Copper homeostasis	876	629	742	1097	752	888
	Zinc resistance	331	130	51	266	143	220
	Arsenic resistance	231	167	144	260	187	218
	Copper homeostasis: copper tolerance	71	90	123	152	119	134
	Resistance to chromium compounds	60	70	74	138	100	135
	Mercury resistance operon	56	56	130	61	64	88
	Mercuric reductase	51	51	116	79	54	73
	Cadmium resistance	12	8	11	13	5	7
Multidrug Resistance Efflux Pump	MexA-MexB-OprM Multidrug Efflux System	2	1	0	4	2	3
	MexC-MexD-OprJ Multidrug Efflux System	0	1	0	1	0	0
	MexE-MexF-OprN Multidrug Efflux System	50	51	55	71	33	29
	Multidrug efflux pump in <i>Campylobacter jejuni</i> (CmeABC operon)	130	173	208	231	228	261
	Multidrug Resistance Efflux Pumps	874	740	1002	1128	952	1022
	Multiple Antibiotic Resistance MAR locus	0	0	1	1	0	2

Table 3.3 (cont'd)

	The mdtABCD multidrug resistance cluster	106	112	124	143	140	131
	Beta-lactamase	409	236	264	592	384	399
	Methicillin resistance in Staphylococci	251	148	217	307	216	207
Resistance to Antibiotics	Erythromycin resistance	65	39	54	77	53	68
	Fosfomycin resistance	3	4	6	1	0	1
	Resistance to fluoroquinolones	649	502	613	900	517	554
	Resistance to Vancomycin	22	4	1	17	11	8
	Adaptation to d-cysteine	1	3	0	2	5	3
	Aminoglycoside adenylyltransferases	3	3	3	2	0	0
Resistance to Other Toxic Compounds	BlaR1 Family Regulatory Sensor-transducer Disambiguation	679	386	496	840	546	614
	Lysozyme inhibitors	0	2	0	1	0	1
	Polymyxin Synthetase Gene Cluster in Bacillus	2	0	1	0	1	0
	Bile hydrolysis	22	28	28	51	24	24

REFERENCES

REFERENCES

1. Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J. and Handelsman, J. (2010) Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology* 8(4), 251-259.
2. Auerbach, E.A., Seyfried, E.E. and McMahon, K.D. (2007) Tetracycline resistance genes in activated sludge wastewater treatment plants. *Water research* 41(5), 1143-1151.
3. Baker-Austin, C., Wright, M.S., Stepanauskas, R. and McArthur, J. (2006) Co-selection of antibiotic and metal resistance. *TRENDS in Microbiology* 14(4), 176-182.
4. Baquero, F., Negri, M.-C., Morosini, M.-I. and Blázquez, J. (1998) Antibiotic-selective environments. *Clinical infectious diseases* 27(Supplement 1), S5-S11.
5. Blasco, M.D., Esteve, C. and Alcaide, E. (2008) Multiresistant waterborne pathogens isolated from water reservoirs and cooling systems. *Journal of Applied Microbiology* 105(2), 469-475.
6. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170
7. Borjesson, S., Melin, S., Matussek, A. and Lindgren, P.E. (2009) A seasonal study of the *mecA* gene and *Staphylococcus aureus* including methicillin-resistant *S. aureus* in a municipal wastewater treatment plant. *Water Res* 43(4), 925-932.
8. Brooks, J.B.J., Maxwell, S.M.S., Rensing, C.R.C., Gerba, C.G.C. and Pepper, I.P.I. (2007) Occurrence of antibiotic-resistant bacteria and endotoxin associated with the land application of biosolids. *Can J Microbiol* 53(5), 616-622.
9. Carlet, J., Jarlier, V., Harbarth, S., Voss, A., Goossens, H., & Pittet, D. (2012). Ready for a world without antibiotics? The penicillins antibiotic resistance call to action. *Antimicrobial resistance and infection control*, 1(1), 1-13.
10. Chandra, R. and Sankhwar, M. (2011) Influence of lignin, pentachlorophenol and heavy metal on antibiotic resistance of pathogenic bacteria isolated from pulp paper mill effluent contaminated river water. *Journal of environmental biology/Academy of Environmental Biology, India* 32(6), 739.
11. Deredjian, A., Colinon, C., Brothier, E., Favre-Bonté, S., Cournoyer, B. and Nazaret, S. (2011) Antibiotic and metal resistance among hospital and outdoor strains of *Pseudomonas aeruginosa*. *Research in microbiology* 162(7), 689-700.
12. Dodd, M. C. (2012). Potential impacts of disinfection processes on elimination and deactivation of antibiotic resistance genes during water and wastewater treatment. *Journal of Environmental Monitoring*, 14(7), 1754-1771.

13. Durso, L. M., Harhay, G. P., Bono, J. L., & Smith, T. P. (2011). Virulence-associated and antibiotic resistance genes of microbial populations in cattle feces analyzed using a metagenomic approach. *Journal of microbiological methods*, 84(2), 278-282.
14. Fahrenfeld, N., Ma, Y., O'Brien, M., & Pruden, A. (2013). Reclaimed water as a reservoir of antibiotic resistance genes: distribution system and irrigation implications. *Frontiers in microbiology*, 4.
15. Fernández, L., & Hancock, R. E. (2012). Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clinical microbiology reviews*, 25(4), 661-681.
16. Francisco, R., Alpoim, M. C., & Morais, P. V. (2002). Diversity of chromium-resistant and-reducing bacteria in a chromium-contaminated activated sludge. *Journal of Applied Microbiology*, 92(5), 837-843.
17. Guardabassi, L., Lo Fo Wong, D. and Dalsgaard, A. (2002) The effects of tertiary wastewater treatment on the prevalence of antimicrobial resistant bacteria. *Water research* 36(8), 1955-1964.
18. Ishaque, A. B., Johnson, L., Gerald, T., Boucaud, D., Okoh, J., & Tchounwou, P. B. (2006). Assessment of individual and combined toxicities of four non-essential metals (As, Cd, Hg and Pb) in the microtox assay. *International journal of environmental research and public health*, 3(1), 118-120.
19. Kamala-Kannan, S. and Lee, K.J. (2008). Metal tolerance and antibiotic resistance of *Bacillus* species isolated from Suncheon Bay sediments, South Korea. *Biotechnology* 7(1), 149-152.
20. Kim, S. and Aga, D.S. (2007) Potential ecological and human health impacts of antibiotics and antibiotic-resistant bacteria from wastewater treatment plants. *Journal of Toxicology and Environmental Health, Part B* 10(8), 559-573.
21. Knapp, C. W., McCluskey, S. M., Singh, B. K., Campbell, C. D., Hudson, G., & Graham, D. W. (2011). Antibiotic resistance gene abundances correlate with metal and geochemical conditions in archived Scottish soils. *PLoS One*, 6(11), e27300.
22. Knapp, C.W., Dolfing, J., Ehlert, P.A. and Graham, D.W. (2009). Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environmental science & technology* 44(2), 580-587.
23. Krasnići, N., Dragun, Z., Erk, M., & Raspor, B. (2013). Distribution of selected essential (Co, Cu, Fe, Mn, Mo, Se, and Zn) and nonessential (Cd, Pb) trace elements among protein fractions from hepatic cytosol of European chub (*Squalius cephalus* L.). *Environmental Science and Pollution Research*, 20(4), 2340-2351.
24. La Scola, B., Elkarkouri, K., Li, W., Wahab, T., Fournous, G., Rolain, J.-M., Biswas, S., Drancourt, M., Robert, C. and Audic, S. (2008). Rapid comparative genomic analysis for clinical microbiology: the *Francisella tularensis* paradigm. *Genome research* 18(5), 742-750.

25. LaPara, T. M., Burch, T. R., McNamara, P. J., Tan, D. T., Yan, M., & Eichmiller, J. J. (2011). Tertiary-treated municipal wastewater is a significant point source of antibiotic resistance genes into Duluth-Superior Harbor. *Environmental science & technology*, 45(22), 9543-9549.
26. Luo, Y., Yang, F., Mathieu, J., Mao, D., Wang, Q., & Alvarez, P. J. J. (2013). Proliferation of Multidrug-Resistant New Delhi Metallo- β -lactamase Genes in Municipal Wastewater Treatment Plants in Northern China. *Environmental Science & Technology Letters*, 1(1), 26-30.
27. Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402.
28. McKinney, C. W., & Pruden, A. (2012). Ultraviolet disinfection of antibiotic resistant bacteria and their antibiotic resistance genes in water and wastewater. *Environmental science & technology*, 46(24), 13393-13400.
29. Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R. and Wilke, A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* 9(1), 386.
30. Munir, M., & Xagorarakis, I. (2011). Levels of antibiotic resistance genes in manure, biosolids, and fertilized soil. *Journal of environmental quality*, 40(1), 248-255.
31. Munir, M., Wong, K. and Xagorarakis, I. (2011.) Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. *Water Res* 45(2), 681-693.
32. Nagulapally, S.R., Ahmad, A., Henry, A., Marchin, G.L., Zurek, L. and Bhandari, A. (2009). Occurrence of ciprofloxacin-, trimethoprim-sulfamethoxazole-, and vancomycin-resistant bacteria in a municipal wastewater treatment plant. *Water Environment Research* 81(1), 82-90.
33. Odjadjare, E. E., Igbinsosa, E. O., Mordi, R., Igere, B., Igeleke, C. L., & Okoh, A. I. (2012). Prevalence of Multiple Antibiotics Resistant (MAR) *Pseudomonas* Species in the Final Effluents of Three Municipal Wastewater Treatment Facilities in South Africa. *International journal of environmental research and public health*, 9(6), 2092-2107.
34. Peltier, E., Vincent, J., Finn, C., & Graham, D. W. (2010). Zinc-induced antibiotic resistance in activated sludge bioreactors. *water research*, 44(13), 3829-3836.
35. Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420-1428.

36. Pruden, A., Pei, R., Storteboom, H. and Carlson, K.H. (2006). Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environmental science & technology* 40(23), 7445-7450.
37. Reinthaler, F., Posch, J., Feierl, G., Wüst, G., Haas, D., Ruckebauer, G., Mascher, F. and Marth, E. (2003). Antibiotic resistance of *E. coli* in sewage and sludge. *Water research* 37(8), 1685-1690.
38. Shi, P., Jia, S., Zhang, X. X., Zhang, T., Cheng, S., & Li, A. (2013). Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water research*, 47(1), 111-120.
39. Sinha, S. (2013). Studies on heavy metal tolerance and antibiotic resistance patterns of bacterial population isolated from effluent treated water of Delhi. *Journal of Biomedical and Pharmaceutical Research*, 2(5).
40. Spain, A. and Alm, E. (2003) Implications of microbial heavy metal tolerance in the environment. *Reviews in undergraduate research* 2(1), 6.
41. Stepanauskas, R., Glenn, T.C., Jagoe, C.H., Tuckfield, R.C., Lindell, A.H. and McArthur, J. (2005) Elevated microbial tolerance to metals and antibiotics in metal-contaminated industrial environments. *Environmental science & technology* 39(10), 3671-3678.
42. to determine antibiotic resistance genotypes. *BMC bioinformatics* 11(Suppl 4), P16.
43. Tuckfield, R.C. and McArthur, J.V. (2008) Spatial analysis of antibiotic resistance along metal contaminated streams. *Microbial ecology* 55(4), 595-607.
44. Volkmann, H., Schwartz, T., Bischoff, P., Kirchen, S. and Obst, U. (2004) Detection of clinically relevant antibiotic-resistance genes in municipal wastewater using real-time PCR (TaqMan). *Journal of Microbiological Methods* 56(2), 277-286.
45. Webber, M. A., & Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, 51(1), 9-11.
46. Wilke, A., Glass, E. M., Bischof, J., Braithwaite, D., D'Souza, M., Gerlach, W., ... & Meyer, F. (2014). *MG-RAST Manual for version 3.3*. 6, revision 6.
47. Wright, M.S., Peltier, G.L., Stepanauskas, R. and McArthur, J.V. (2006) Bacterial tolerances to metals and antibiotics in metal-contaminated and reference streams. *FEMS microbiology ecology* 58(2), 293-302.
48. Xi, C., Zhang, Y., Marrs, C. F., Ye, W., Simon, C., Foxman, B., & Nriagu, J. (2009). Prevalence of antibiotic resistance in drinking water treatment and distribution systems. *Applied and environmental microbiology*, 75(17), 5714-5718.

49. 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), e26041.
50. Zhang, X.-X. and Zhang, T. (2011) Occurrence, abundance, and diversity of tetracycline resistance genes in 15 sewage treatment plants across China and other global locations. Environmental science & technology 45(7), 2598-2604.
51. Zhang, Y., Marrs, C.F., Simon, C. and Xi, C. (2009). Wastewater treatment contributes to selective increase of antibiotic resistance among *Acinetobacter* spp. Science of the Total Environment 407(12), 3702-3706.

CHAPTER 4

SCREENING FOR POTENTIAL VIRAL AND BACTERIAL PATHOGENS IN WASTEWATER EFFLUENT RELEASED FROM AN MBR AND A CONVENTIONAL TREATMENT UTILITY USING METAGENOMICS ANALYSIS

Mariya Munir, Terence Marsh, and Irene Xagorarakis. (in preparation)

Abstract

Despite recent rapid advancements in water and wastewater treatment technologies, waterborne pathogens still remain as one of the major environmental threats to human health. Monitoring of all pathogens with conventional methods is not feasible due to time and cost constraints. In this paper, virus and bacterial diversity of two wastewater treatment plants, a conventional activated sludge and membrane bioreactor (MBR), are investigated using metagenomics. Effluent samples (before and after disinfection effluents) have been analyzed to reveal microbial diversity. Diversity analysis does not provide quantitative data on pathogen loads or infectivity but it provides a list of potentially pathogenic viruses and bacteria that need to be considered in more detail. *Caudovirales* is the most dominating order in the viral community detected in our samples, consisting of families: *Siphoviridae*, *Myoviridae* and *Podoviridae*. Analyzing the bacterial community, Proteobacteria was the highly abundant bacterial phylum followed by *Bacteroidetes*, *Planctomycetes*, *Actinobacteria* and *Firmicutes*.

Further, the most abundant potential human viral pathogen observed in our study belongs to taxonomic order *Herpesviridales*. Other potentially pathogenic viruses detected in this study include *Adenoviridae* and *Coronaviridae*. While all the bacterial pathogens described in the Contaminant Candidate List from EPA were detected in our study which include *Cyanobacteria*,

Legionella pneumophila, *Salmonella enterica* and *Aeromonas hydrophila*, *Mycobacterium avium intracellulare* (MAC), *Helicobacter pylori*, *Campylobacter jejuni*, *Shigella sonnei* and *Escherichia coli* (O157). Other common bacterial pathogens found in our samples are *Leptospira*, *Mycobacterium tuberculosis*, *Vibrio cholera*, *Yersinia pestis*, and *Yersinia enterocolitica*. Metagenomic analyses in this study also revealed that a large proportion of sequences could not be assigned to taxonomic affiliations even at the phylum/class levels and thus are most likely to be derived from novel, uncharacterized microbes. This study provides a complete description of virus and bacterial diversity in effluents from a conventional utility and an MBR utility. This paper provides guidance on which pathogens to monitor in the effluents and suggest WWTPS as are reservoirs of microbial populations of public health relevance

Keywords: Metagenomics, wastewater, pathogenic virus, pathogenic bacteria, Illumina Hiseq

1. Introduction

Despite recent advances in water and wastewater treatment technology, waterborne diseases still pose a serious threat to public health across the world. List of contaminants (called the Contaminant Candidate List or CCL) generated by US EPA (Environmental Protection Agency) are the contaminants that are known or anticipated to occur in public water systems, and which may require regulation under the Safe Drinking Water Act (SDWA). Table 4.1 describes a list of microbes on CCL. Table 4.2 and 4.3 provide some other common pathogens detected in effluents and raw samples respectively from WWTPs.

Viruses are potentially the most important and most hazardous among the pathogens found in wastewater (Sidhu et al., 2008, Toze, 1997). They are also generally more difficult to detect in environmental samples. A high diversity of human viral pathogens is present in our

environment (approximately 200 recognized viral pathogen species) which are further elevated in environments affected by pollution, and additional species are continuously discovered every year (Bibby, 2013). The most commonly detected pathogenic viruses are the enteroviruses. Other viruses which have been detected in wastewater include adenoviruses, rotaviruses, reoviruses, astroviruses and caliciviruses (such as Norwalk virus). Bacteria is the most common type of microorganism found in wastewater and bacterial indicators like *E.coli* and fecal coliforms are the required microbial measures of effluents for wastewater-discharge (Francy et al. 2011). The list of pathogenic bacteria is quite large and can be further classified as waterborne faecal pathogens and non-faecal pathogens. Waterborne faecal pathogens are microorganisms that are result of contamination from human or animal faeces and are often associated with bacteria from the Enterobacteriaceae family. They are well established as having a history of being responsible for waterborne outbreaks of gastrointestinal illness and include pathogenic *Escherichia coli*, *Salmonella*, *Shigella*, *Campylobacter* and *Yersinia*. Non-faecal pathogens are usually bacteria naturally found in source waters and have a public health impact. They include *Legionella*, *Mycobacterium avium* complex, *Aeromonas* and *Helicobacter pylori* (Health Canada, 2013).

Even state of the art WWTPs such as MBRs (Membrane Biological Reactor) have been proven to release pathogenic viruses in the environment (Simmons and Xagorarakis, 2011). Conventional utilities release pathogenic viruses (Simmons and Xagorarakis, 2011) in the effluent and potentially pathogenic viruses in sludge (Bibby and Peccia, 2013). Similarly, potentially pathogenic bacteria were also detected in the activated sludge and effluents from WWTPs (Ye and Zhang, 2011, Odjadjare, 2010). According to a recent study, human enteric viruses were detected in effluent from two different WWTPs using RT-QPCR (Kitajima et al. 2014). Another study based on characterizing effluent water quality from satellite MBRs facilities reported that

adenoviruses were detected in effluent from all nine MBR facilities sampled (Hirani et al. 2013). There have been reports of finding pathogens in the effluent from different WWTPs even after disinfection treatment (Kitajima et al. 2014, Hirani et al. 2013, Simmons et al. 2011, Fong et al. 2010, Okoh et al. 2007). Further research is required to identify a complete list of potentially infectious viruses and bacteria in the effluent of WWTPs.

Next-generation DNA sequencing has recently been applied to study viral metagenomes (viromes) and bacterial diversity in different environmental samples (Aw et al. 2014; Alhamlan et al. 2013, Gomez-Alvarez et al. 2012, Hu et al. 2012, Bibby et al. 2011, Wommack et al. 2011, Tamaki et al. 2011, Rosario et al. 2009). The emergence of next-generation sequencing as a new research tool has aided in the detection of genetically diverse and rapidly evolving novel pathogens (Malik and Matthijnssens, 2014; Naccache et al. 2014). With the help of metagenomic tools, microbial communities related with wastewater could easily be analyzed. According to Zhang et al (2011), a comprehensive characterization of the vast microbial community present in activated sludge systems is hindered by the low sequencing depth of the traditional PCR-cloning approach (Zhang et al. 2011). Previous methods, however, were limited by a requirement that a researcher must select the microbes (pathogenic viruses and bacteria) that will be searched for, but in contrast, the metagenomic analyses produces a list of microbes that is based on abundance and is independent of researcher bias (Bibby et al. 2011). Further, it has been found that many species are difficult to isolate because they fail to grow in laboratory culture, depend on other organisms for critical processes, or have become extinct. Next-generation sequencing methods overcome these obstacles, as DNA can be isolated directly from living or dead cells in various contexts (Tringe et al. 2005). It is expected that these powerful new methods will open up new questions to genomic exploration and will also allow high-throughput sequencing to be more

than just a discovery exercise but also a routine assay for hypothesis testing. However, these methods also have certain limitations. They can't provide infectivity results, are good for family level identification and identification is as good as the assembly methods and database used.

The goal of this study was to describe and compare the overall diversity of viruses and bacteria and screen for potentially pathogenic viruses and bacteria present in a conventional and MBR wastewater treatment plant effluents using metagenomic investigations. Metagenomic sequencing has been adapted for detection and characterization of viruses in complex wastewater environments (Aw et al., 2014). This whole genome sequencing approach can be considered as good as database comparison and is expected to be more than just a discovery exercise and also a routine assay for hypothesis testing. Usage of assembled sequence data (contigs) for analysis reduces the chances of false positive detection in our study.

2. Materials and Methods

2.1. Sample Collection: Effluent samples before and after disinfection were collected from East Lansing (EL) WWTP and Traverse City (TC) WWTP in Michigan (U.S.A.) in 2013. Table 4.4 provides the characteristics of these WWTPs based on wastewater treatment processes and disinfection methods. Two liters of grab effluent samples were collected for bacterial isolation two samples from each WWTP while for viral isolation, Argonite filters were used. Approximately 400 Litres of effluent samples were passed through the filter pumped through a sampler at a rate of about 11-12L/min. Samples were kept on ice and were transported to the Water Quality Engineering Laboratory at Michigan State University (East Lansing, U.S.A.) for further immediate processing.

2.2. Sample Processing:

Bacteria in the effluent samples were concentrated by filtration with 0.45 μm HA filters (Millipore, Billerica, MA). The volume of effluent samples filtered was 1 L for each of the four samples. The filters were collected in a 50 mL tubes and 50 mL Phosphate Buffer Water (PBW) was added in each tube containing a filter. The tubes were then vortexed for 5 min to allow the biomass layer on the filters to mix with water. All the tubes were then centrifuged for 20 min at 4500rpm to concentrate the sample down to 2 mL. Supernatant was discarded and the concentrates were stored at $-80\text{ }^{\circ}\text{C}$ until the DNA extraction was performed for further molecular analysis.

All virus samples collected were eluted 12–24 h after initial sampling according to the Concentration and Processing of Waterborne Viruses by Positive Charge 1MDS Cartridge Filters and Organic Flocculation (USEPA, 2001). Briefly, a 1.5% w/v beef extract (0.05 M glycine, pH 9.0–9.5) solution was used for eluting the filters. The filters were submerged for a total of 2 min (2 separate 1 min elutions) in filter housings with 1 L of beef extract added to the pressure vessel. The filter housing was disinfected between filters using 0.17% bleach solution for a 1 min contact time and then dechlorinated using 2% sodium thiosulfate for another 1 min contact time. After the beef extract was passed through each filter, the 1 L of beef extract and eluted particles had the pH adjusted to 3.5 ± 0.1 using 1 M HCl and slowly flocculated for 30 min. Further concentrated of the solution was done by placing 500 mL into a centrifuge bottle centrifuge for 15 min at $2500\times g$ at $4\text{ }^{\circ}\text{C}$. The supernatant was then slowly poured off and the process was repeated until all the beef extract solution was centrifuged. The accumulated pellets were resuspended using 30 mL of 0.15 M sodium phosphate (pH 9.0–9.5), mixed until the pellet was mostly dissolved and the pH was adjusted to 9.0–9.5 using 1 M HCl. Next, the solution was placed into a 40 mL centrifuge tube and centrifuge for 10 min at $4\text{ }^{\circ}\text{C}$ at $7000\times g$. The supernatant

was poured off into a 50 mL centrifuge bottle, the pH was adjusted to 7.0–7.5 for stabilization of the virus particles and the pellet was discarded. The supernatant was loaded into a 60 mL syringe and passed through a 0.22 µm sterilized filter for removal of bacteria, fungi and other contaminating agents. All samples were completely mixed and placed into 2 mL cryogenic tubes and stored at –80 °C until further analysis.

2.3. Nucleic acid Extraction: Bacterial DNA and virus DNA was extracted (from samples before and after disinfection) using a MagNA Pure Compact DNA extractor (Roche Applied Science, Indianapolis, IN, USA) following the protocol in the manufacturer’s manual. The MagNA Pure Compact utilizes a magnetic-bead technology for the isolation process. Sample amount of 400 µL was loaded in the system and the elution volume was 100µL. The extracts were stored in a freezer at -20°C. Following extraction the quantity of bacterial and viral nucleic acid extracts from all samples were checked using the NanoDrop Spectrophotometer (NanoDrop® ND-1000, Wilmington, DE).

2.4. Metagenomic analyses: Bacterial and Virus-enriched DNA isolated from the effluent samples was sequenced on an Illumina platform (Illumina HiSeq, Roche Technologies) at The Research Technology Support Facility (RTSF) genomic center at Michigan State University generating 250 bp paired-end reads. A flexible read trimming tool for Illumina NGS data called Trimmomatic was used for trimming Illumina data and removing adapters (Bolger et al. 2014). Running Trimmomatic is a good first step in quality filtering the Illumina data. The sequences were assembled using an iterative De Bruijn Graph De Novo Assembler for Short Reads Sequencing data with Highly Uneven Sequencing Depth called **IDBA-UD** (Peng et al. 2012). Once the sequences were assembled into contigs, diversity analysis was done for bacteria and viruses using MGRAST and METAVIR analysis platform respectively. After the diversity

analysis done, search was done for CCL pathogens listed in Table 4.1 along with some other emerging pathogens.

2.4.1. MG-RAST analyses: The assembled data from all the samples were analyzed using MG-RAST 3.3.1 (Meyer et al. 2008). Data was analyzed based on organism abundance and on the functional distribution at the subsystem hierarchy with maximum E- value cutoff of $1E^{-5}$, minimum percent identity cutoff of 60% and minimum alignment length cutoff of 15bps. The MG-RAST pipeline is a fully automated pipeline providing quality control, feature prediction and functional annotation (Thomas et al. 2012). Analysis includes the phylogenetic comparisons and functional annotations against the database and the results are expressed in the form of abundance profiles. MG-RAST searches the non-redundant M5NR and M5RNA databases in which each sequence is unique. The best hit classification option was used which reports the functional and taxonomic annotation of the best hit in the M5NR for each feature. The number of hits is defined as “occurrences of the input sequence in the database”. The “abundance” in the MG-RAST presents the estimate of the number of sequences that contain a given annotation, found by multiplying each selected database match (hit) by the number of representatives in each cluster. Hits refer to the number of unique database sequences that were found in the similarity search, not the number of reads (Wilke et al. 2014). Some of the displayed data has been normalized to values between 0 and 1 to allow for comparison of differently sized samples. Each of the categories was further studied for detailed analysis and data was downloaded in Microsoft Office Excel. MG-RAST version 3.0 works with reads of 75bp or longer.

2.4.2. METAVIR 2 analyses: The assembled data from the six virus samples were analyzed using METAVIR 2 (Roux et al. 2014). METAVIR 2 is a web server designed to annotate viral metagenomic sequences (raw reads or assembled contigs). The MetaVir server uses the

taxonomic composition tool, which gives taxonomic affiliations of the viral sequence reads. Taxonomic composition is computed from a BLAST comparison with the Refseq complete viral genomes protein sequences database from NCBI (release of 2014-03-13) using BLASTp. Maximum E- value cutoff of $1E^{-5}$ was used. Predicted proteins from contigs are compared to Refseq through BLAST, and then each protein is affiliated to its best BLAST hit (if any). Then a contig is affiliated to its best BLAST hit, i.e. to the affiliation of the predicted protein with the higher BLAST score. The number of hits is defined as “occurrences of the input sequence in the database”. Best hit ratio is defined as the number of hits for one category divided by total number of hits. Metavir only select for sequences longer than 300 bp.

3. Results and Discussion:

3.1. Metagenome sequencing and assembly results:

To describe the microbial community in wastewater treatment plant (before and after disinfection effluents), Illumina Hiseq sequencing was applied on eight samples from two different WWTPs in Michigan. Virus and bacterial DNA was isolated from two different locations (before disinfection effluent (BD) and after disinfection effluent (AD)) from a conventional and MBR utility. Sequences generated were assembled using an IDBA-UD assembler. Virus metagenomes were analysed using METAVIR 2 software tool while bacterial metagenomes were analysed using MGRAST v3 analysis tool. Metagenomic analysis details for virus isolated samples are presented in Table 5. As computed by METAVIR, virus isolated DNA samples contained contigs ranging from 151992 to 256064 sequences. Virus isolated effluent sample after disinfection from EL contains 11.6% of affiliated sequences, while before disinfection effluent showed 19.39% of affiliated sequences. In virus isolated effluent sample

after disinfection from TC, 16.36 % of the total sequences are affiliated while before disinfection sample contains 19.85 % of affiliated sequences.

Metagenome analysis revealed bacterial DNA samples contained contigs ranging from 238657 to 288930 sequences. In bacterial isolated effluent sample after disinfection from EL, 4,302 sequences (1.8%) failed to pass the Quality Control (QC) pipeline. Of the sequences that passed QC, 841 sequences (0.4%) contain ribosomal RNA genes as computed by MGRAST analysis software. Of the remainder, 184,215 sequences (77.2%) contain predicted proteins with known functions and 48,954 sequences (20.5%) contain predicted proteins with unknown function. 339 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Out the 234,355 sequences (totaling 147,771,487 bps) that passed quality control, 233,169 (99.5%) produced a total of 270,966 predicted protein coding regions. Of these 270,966 predicted protein features, 172,766 (63.8% of features) have been assigned an annotation using at least one of the protein databases (M5NR). 148,347 features (85.9% of annotated features) were assigned to functional categories. Similarly, in before disinfection effluent samples, 5,765 sequences (2.2%) failed to pass the QC pipeline. Of the sequences that passed QC, 702 sequences (0.3%) contain ribosomal RNA genes. Of the remainder, 176,685 sequences (68.8%) contain predicted proteins with known functions and 70,878 sequences (27.6%) contain predicted proteins with unknown function. 2,845 (1.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Of the 251,116 sequences (totaling 153,515,277 bps) that passed quality control, 247,563 (98.6%) produced a total of 297,712 predicted protein coding regions. Of these 297,712 predicted protein features, 167,031 (56.1% of features) have been assigned an annotation using at least one of the protein databases (M5NR). 137,654 features (82.4% of annotated features) were assigned to functional categories.

In bacterial isolated effluent sample after disinfection from TC, 3,695 sequences (1.3%) failed to pass the QC pipeline. Of the sequences that passed QC, 624 sequences (0.2%) contain ribosomal RNA genes. Of the remainder, 231,252 sequences (80.0%) contain predicted proteins with known functions and 53,202 sequences (18.4%) contain predicted proteins with unknown function. 153 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. While in before disinfection effluent samples, 6,703 sequences (2.6%) failed to pass the QC pipeline. Of the sequences that passed QC, 552 sequences (0.2%) contain ribosomal RNA genes. Of the remainder, 202,533 sequences (78.1%) contain predicted proteins with known functions and 49,351 sequences (19.0%) contain predicted proteins with unknown function. 77 (0.0%) of the sequences that passed QC have no rRNA genes or predicted proteins. Analysis statistics for all the bacterial isolated sequenced samples are shown in Table 4.6 (a &b).

3.2. Microbial diversity in WWTP:

The diversity of viruses and bacteria present in effluent samples from a full-scale conventional activated sludge wastewater treatment plant and a membrane bioreactor (MBR) utility was studied using Illumina Hiseq sequencing. Analysis software METAVIR 2 and MGRAST v3.0 pipeline was used for analyzing the assembled sequences. Virus and bacterial species richness is presented in the rarefaction curve in Figure 4.3. The rarefaction curve of annotated species richness is a plot of the total number of distinct species annotations as a function of the number of sequences sampled (Meyer et al. 2008). These rarefaction curves are calculated from the table of species abundance. Rarefaction curve also demonstrate whether a sample has been sequenced to saturation. The slope of the right-hand part of the curve is related to the fraction of sampled species that are rare (Wilke et al. 2013). A steeper slope indicates that

the sample has not yet been fully sequenced (that is large fraction of the species diversity remains to be discovered. Figure 4.3 shows that diversity of bacteria is higher from a conventional (EL WWTP) compared to MBR utility (TC WWTP) in effluent. This is as expected because MBR utilities are said to provide better treatment comparatively. Also, bacterial and virus diversity is higher in samples from before disinfection compared to after disinfection which is how the situation should be. Based on alpha diversity (Figure 4.3), virus diversity is higher in conventional (EL WWTP) compared to MBR utility (TC WWTP) in effluent samples. The alpha diversity estimate summarizes the distribution of species-level annotations in each metagenome dataset; it is the diversity of organisms in a sample with a single number. In simpler word, it is used to describe the number of distinct species in a given sample.

3.2.1. Virus diversity in WWTP:

Viral diversity was explored by Illumina Hiseq sequencing on virus-enriched DNA obtained from effluent samples. Table 4.7 provides the taxonomic comparison of viruses based on Baltimore classification. This classification groups viruses into families, depending on their type of genome (DNA, RNA, single-stranded (ss), double-stranded (ds), etc.) and their method of replication. Most of the viruses found in all our samples belong to group of dsDNA viruses with no RNA stage. Table 7 shows the viral taxonomic composition of the all the assembled datasets from the wastewater viromes obtained from METAVIR server. The taxonomic compositions are computed for each dataset using the BLASTp based on best BLAST hit affiliation of each contigs (threshold of 50 on the BLAST bitscore). The best hit ratios are calculated based on all hits to Refseq Virus, and is calculated from dividing the number of hits from one category (e.g. dsDNA viruses) by the total number of hits. In order to relate these ratios to the complete metagenome they can be further normalized with ratio of affiliated sequences.

Virus families are classified according to the NCBI taxonomy; each column represents a dataset and each row a group of virus families. From the results we found that *Caudovirales* is the most dominating order accounting for 56.10% - 78.11% of total virus sequences. Among this order, *Siphoviridae* is the most dominating virus family closely followed by *Myoviridae* and *Podoviridae*. These three families belong to order Caudovirales which are dsDNA group are also known as tailed bacteriophages. Phycodnaviridae and Mimiviridae are other two families of viruses that are present in high numbers in all the samples. Most of the sequence reads were unclassified that must be derived from putatively novel viruses. There are more than 3,000 different viruses are recognized, but it has been suggested using metagenomic studies that these are a small fraction of the total that exist in nature (Cantalupo et al. 2011). It is very important to identify the virus diversity as the lack of knowledge on characteristics of the viral universe and the diversity of viral genomes is a roadblock to understanding important issues, like the origin of emerging pathogens and the extent of gene exchange among viruses (Cantalupo et al. 2011). Most sequences showed no sequence relation to any known sequences in the databases and thus are most likely to be derived from novel, uncharacterized viruses.

3.2.2. Bacterial diversity in WWTP:

Diversity of bacteria in effluent samples was explored at the phylum taxonomic level using the MG-RAST server. Table 4.8 shows the organism abundance based in best hit classification. The displayed data was compared to M5NR using a maximum e-value of $1e-5$, a minimum identity of 60 %, and a minimum alignment length of 15 and has been normalized to values between 0 and 1 to allow for comparison of differently sized samples. Normalization allows for correction based on the sample size. *Proteobacteria* was the most abundant phylum in all samples, accounting for (71-81%), of total effective bacterial sequences. This is consistent

with the results of bacterial communities found in sewage (Zhang et al., 2011; McLellan *et al.*, 2010) and soil (Roesch *et al.*, 2007). This phylum was followed by *Bacteroidetes* accounting (8-14%) in phylum distribution in all the samples. The other dominant phyla were *Planctomycetes* (1-3.4%), *Actinobacteria* (1.2–2.8%), *Firmicutes* (1.6-2.1%) and *Verrucomicrobia* (1-3%). This is very similar to few other studies on wastewater and specially sludge (Zhang et al. 2011; Xia et al., 2010; Snaidr et al., 1997). Apart from these four dominating groups in this study, a few other major (average abundance >1%) phyla includes *Chloroflexi*, *Cynobacteria*, *Nitrospira* and *Acidobacteria*. The four classes of the *Proteobacteria* abundant in all the samples are *Betaproteobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria* and *Deltaproteobacteria*. Metagenomic analysis showed that the bacterial community composition of effluent samples was distinct between the two wastewater treatment systems.

Difference in the abundance hit with respect to each sample is observed mostly because of different sample size (sequence length), so the diversity comparison among different sample is not possible from this data. Bacterial diversity was similar in EL WWTP and TC WWTP in all the samples but some genus was more abundant in ELWWTP and some in TC WWTP, thus a slight variation in population is observed.

3.3. Potential Pathogens in WWTP:

In order to explore pathogenic viruses and bacteria in wastewater, Illumina Hiseq sequencing has proven to be an effective tool. Among the groups of pathogenic viruses tested based on Table 4.1, four different pathogenic viruses were detected in our samples. Pathogenic bacteria abundance based on CCL and other common potential pathogens is presented in Figure 4.4. The most abundant potential human pathogenic virus observed in our study belongs to

taxonomic order *Herpesviridales* (Table 4.7). *Herpesviridae* is a large family of DNA viruses that cause diseases in humans and animals. Herpes viruses have been regarded as the leading cause of human viral disease, along with influenza and cold viruses and it is said that once a patient has become infected by herpes virus, the infection remains for life (Hunt, 2011). More than eight types of Herpes viruses are known to infect humans including Herpes simplex virus Type 1 (HSV-1), Herpes simplex virus Type 2 (HSV-2), Human herpes virus 3 (Varicella Zoster Virus (VZV)), Human herpes virus 4 (Epstein Barr virus (EBV)), Human herpes virus 5 (Cytomegalovirus (CMV)), Human herpes virus 6/7 (exanthum subitum or roseola infantum), Human herpes virus 8 (Kaposi's sarcoma-associate herpes virus) (Hunt, 2011). All of these human herpes viruses except Human herpes virus 5 were detected in our samples. Apart from Human *Herpesvirus* (1, 2, 3, 4, 6B, 7and 8), several other bovine *Herpesviruses* were also identified.

The other pathogenic viruses detected were Adenoviridae, followed by Coronaviridae. Adenoviridae are the dsDNA viruses having no RNA stage. They usually cause respiratory infections, but they can also cause conjunctivitis, gastroenteritis, cystitis, and rash illness The family Adenoviridae consists of five genera among which Aviadenoviruses (which infects birds) and Mastoadenoviruses (which infect mammals) are present in our samples. Coronaviruses belong to the group of single strand (ss) positive strand viruses having no DNA stage. Coronaviruses primarily infect the upper respiratory and gastrointestinal tract and are the cause all common colds in human adults. These results are consistent with previous studies as shown in Table 4.2. Our results closely matches with studies conducted on sewage sludge (Bibby and Peccia, 2013a, Bibby and Peccia, 2013b). It is possible that the presence and abundance of viral pathogens in wastewater will also vary seasonally. And therefore it is recommended that samples

should be collected and analyzed throughout the year to compare seasonal occurrence of pathogenic viruses and the changing virome (Aw et al., 2014; Katayama et al., 2008). Occurrences of human adenoviruses in wastewater treatment plants and drinking water treatment plants have previously been confirmed by QPCR methods (La Rosa et al. 2010; Albinana-Gimenez et al. 2009).

Pathogenic bacteria abundance based on CCL and other common potential pathogens is presented in Figure 4.5. All the pathogens presented in CCL as mentioned in Table 1 are detected in all our effluent samples. *Cyanobacteria* are the most abundant member from the CCL present in all the samples. They occur worldwide and some species of *cyanobacteria* produce toxins that affect animals and humans. Exposure to cyanobacterial toxins can occur by number of ways (drinking or bathing in contaminated water or recreational water contact). A gram-negative bacteria belonging to genus *Legionella* named as *Legionella pneumophila* is an important human pathogen present in all our samples with highest abundance. This was followed by other CCL pathogens like *Mycobacterium avium*, *Salmonella enterica* and *Aeromonas hydrophila*, *Mycobacterium avium intracellulare* (MAC), *Helicobacter pylori*, *Campylobacter jejuni*, *Shigella sonnei* and *Escherichia coli* (O157) (figure 4.5). Other common bacterial pathogens found in our samples are *Leptospira*, *Mycobacterium tuberculosis*, *Vibrio cholera*, *Yersinia pestis*, and *Yersinia enterocolitica*. These results are consistent with previous studies as shown in Table 4.2.

Metagenomic analyses of wastewater samples in this study also revealed that a large proportion of sequences could not be assigned to taxonomic affiliations even at the phylum/class levels. The minority populations that are hard to be explored by traditional molecular methods can easily be detected using metagenomic analyses. The effluent quality varied greatly due to differences in sewage composition, organic loading, pH, temperature, dissolved oxygen, sludge

retention time applied at the aeration tank and different disinfection methods. There is a possibility of difference in the organism load in the human population for each viruses and bacteria and that the occurrences of these microbes fluctuate daily and seasonally in raw sewage. Samples analyzed in this study represent only a single time point at each WWTP and the diversity could vary if sampling is done in different seasons. Further, it is important to acknowledge the limitations of this research in terms of available resources. Still, this is first of the study to provide complete description of virus and bacterial diversity in effluents from two different wastewater treatment plants. More comparative studies are needed based on well-designed sampling plans to find out the main factors influencing the microbial communities.

4. Conclusions:

In the present work, we analyzed effluent samples using metagenomic analysis in an effort to better understand the composition and diversity of viruses and bacteria in wastewater. The analyses unveiled that the communities are highly diverse. In this paper, we found that *Caudovirales* is the most dominating order in the viral community detected in our samples, consisting of families: *Siphoviridae*, *Myoviridae* and *Podoviridae*. Analyzing the bacterial community, *Proteobacteria* was the highly abundant bacterial phylum followed by *Bacteroidetes*, *Planctomycetes*, *Actinobacteria* and *Firmicutes*.

This study provided a bioinformatics approach for identifying potential microbial pathogens in wastewater. The most abundant potential human pathogen observed in our study belongs to taxonomic order *Herpesviridales*. It is recommended that further study should be conducted to look into *Herpesviruses* since they are present in such a high number. Other pathogenic viruses detected in this study include *Adenoviridae*, and *Coronaviridae*.

While all the bacterial pathogens described in the Contaminant Candidate List from EPA (*Cyanobacteria*, *Legionella pneumophila*, *Mycobacterium avium*, *Salmonella enterica* and *Aeromonas hydrophila*, *Mycobacterium avium intracellulare* (MAC), *Helicobacter pylori*, *Campylobacter jejuni*, *Shigella sonnei* and *Escherichia coli* (O157). Other bacterial pathogens observed in our samples are *Leptospira*, *Mycobacterium tuberculosis*, *Vibrio cholera*, *Yersinia pestis*, and *Yersinia enterocolitica*. As the emphasis of this study was on pathogenic species, the abundance of pathogens found in these samples suggests that these systems are reservoirs of microbial populations of public health relevance. This paper provides guidance on which pathogens to monitor in the effluents. Further, it is very much recommended to test all potential pathogens detected with QPCR or culture methods when wastewater reuse is done example for irrigation purposes etc. This is very important for microbial risk assessment. This study provided an overview of microbial diversity and metagenomics as a basic screening tool for detecting potential pathogens in wastewater environment.

5. Acknowledgement:

We would like to thank the managers of the East Lansing Wastewater Treatment Plant and Traverse City Wastewater Treatment Plant for providing the samples and information needed for this study. A very special thank to Bioinformatic Research Specialist Dr. Tracy Teal. Also, we would like to acknowledge bioinformatic support and assistance provided by Bioinformatic team at High Performance Computing Center (HPCC) at Michigan State University.

APPENDIX

Table 4.1: Microbial Contaminant Candidate List from EPA.

Microbial Contaminant Candidates (Adapted from U.S. EPA 1998, 2005,2009)	Diseases/ Symptoms	CCL-1	CCL-2	CCL-3
VIRUSES				
Adenoviruses	Respiratory illness, and occasionally gastrointestinal illness	✓	✓	✓
Caliciviruses	Gastrointestinal illness (Diarrhea, vomiting, nausea and stomach pain)	✓	✓	✓
Coxsackieviruses	Respiratory illness, meningitis (an infection of the linings of the spinal cord and brain), encephalitis (inflammation of the brain), pleurodynia (chest pain), and myopericarditis (inflammation of the heart)	✓	✓	
Echoviruses	Aseptic meningitis (Pneumonia like symptoms)	✓	✓	
Enterovirus	Respiratory illness			✓
Hepatitis A virus	Liver disease and jaundice			✓
BACTERIA				
<i>Aeromonas hydrophila</i>	Gastroenteritis (inflammation of the stomach and intestines), hemolytic syndrome and kidney disease	✓	✓	
<i>Campylobacter jejuni</i>	Gastrointestinal illness, Guillain-Barré syndrome (disorder affecting peripheral nervous system)			✓
<i>Cyanobacteria</i>	Toxins to cause poisoning (affect the nervous and respiratory systems), liver damage	✓	✓	
<i>Escherichia coli (0157)</i>	Gastrointestinal illness and kidney failure			✓
<i>Helicobacter pylori</i>	Ulcers and stomach cancer	✓	✓	✓
<i>Legionella pneumophila</i>	Lung diseases (severe form of pneumonia)			✓
<i>Mycobacterium avium (MAC)</i>	Lung disease (pulmonary pathogen, tuberculosis and gastrointestinal symptoms)	✓	✓	✓
<i>Salmonella enterica</i>	Salmonellosis (food poisoning), enteric fever, gastroenteritis			✓
<i>Shigella sonnei</i>	Shigellosis (food-borne illness), gastrointestinal illness, bloody diarrhea			✓

Table 4.2: Virus and bacterial Pathogens detected in wastewater effluent.

Pathogens Detected		Methods of detection	Full scale Conventional or MBR	References
Viruses	Bacteria			
Adenoviruses, enterovirus, Norovirus GI		QPCR/RT-QPCR	MBR and conventional	Francy et al. 2012
Adenovirus (HAdV) and Enterovirus (EV)		QPCR	MBR and conventional	Simmons & Xagorarakis, 2011
Adenovirus (HAdV), Enterovirus (EV) and Norovirus		QPCR	MBR and conventional	Simmons et al. 2011
Enterovirus, Norovirus, Adenovirus		QPCR/RT-QPCR	conventional	Hewitt et al. 2011
Human Adenoviruses		QPCR	MBR	Kuo et al. 2010
Enteroviruses		BGM cell culture	conventional	Costán-Longares et al. 2008
Adenovirus, Norovirus GI & GII, enterovirus		QPCR	conventional	La Rosa et al. 2010
Noroviruses		RT-QPCR	MBR and conventional	DaSilva et al. 2007
Enteroviruses	<i>Salmonella typhi, Shigella</i>	QPCR	-	Zhang et al. 2014
Calicivirus (Norovirus and Sapovirus),	<i>Escherichia coli (E. coli)</i>	RT-PCR	MBR	Sima et al. 2011
Enterovirus and Adenovirus	<i>Clostridium perfringens</i>	-	conventional	Jacangelo et al. 2003
Human enteric viruses	<i>C. perfringens, E.coli</i>	culture	conventional	Payment et al. 2001
	<i>Mycobacterium tuberculosis, Bacillus Anthracis, Yersinia pestis and Vibrio cholerae</i>	Illumina MiSeq	conventional	Kumaraswamy et al. 2014
	<i>M.tuberculosis, S. flexneri</i>	Illumina Hiseq 2000	conventional	Cai and Zhang, 2013
	<i>Listeria strains.</i>	culture	conventional	Odjadjare et al. 2010
	<i>Aeromonas hydrophila, B. cereus, C. perfringens, E. faecalis, E. coli, K. pneumoniae, P. aeruginosa, Salmonella spp.</i>	DNA microarray, QPCR,	conventional	Lee et al. 2008

Table 4.2 (cont'd)

	<i>Clostridium perfringens</i> , <i>Escherichia coli</i> , <i>Enterococcus faecalis</i> , <i>Klebsiella pneumoniae</i>	<i>QPCR</i>	conventional	Shannon et al. 2007
	<i>A. hydrophila</i> , <i>B. cereus</i> , <i>C. perfringens</i> , <i>E. faecalis</i> , <i>E. coli</i> , and <i>K. pneumoniae</i> , <i>P. aeruginosa</i> .	DNA microarray, QPCR,	conventional	Lee et al. 2006

Table 4.3: Virus and bacterial Pathogens detected in raw and sludge samples.

Pathogens Detected		Methods of detection	Full scale Conventional or MBR	References
Viruses	Bacteria			
Adenovirus, Aichi virus, Astroviruses, Bocavirus, Coronavirus, Cosavirus, Echovirus, Hepatitis C, Herpesvirus, Human Immuno deficiency virus, Klassevirus, Norovirus, Norwalk virus, Papillomavirus, Parechovirus, Parvovirus, Poliovirus, Rhinovirus, Rotavirus, Rubella virus, T-Lymph virus, Torque Teno Virus		Illumina Hiseq 2000		Bibby et al. 2013
Adenovirus species B and C		Ion Torrent sequencing	conventional	Bibby et al. 2013 b
Adenovirus, Parechovirus, Coronavirus, Torque teno virus, Herpes virus and Aichi		Pyrosequencing	conventional	Bibby et al. 2011
Adenovirus, Norovirus GI & GII, and Enterovirus.		QPCR	conventional	La Rosa et al. 2010
Enterovirus, rotavirus and norovirus	<i>Salmonella and Shigella</i>	QPCR	MBR	Zhou et al. 2014
Adenoviruses, Calciviruses (noroviruses), Enteroviruses, hepatitis A viruses and hepatitis E viruses	<i>Escherichia coli O157:H7, Helicobacter pylori, Legionella pneumophila, Campylobacter jejuni, Mycobacterium avium complex, Salmonella enterica, Shigella spp</i>	QPCR	-	Aw and Rose, 2012
				Yates, 2011
				Griffin et al. 2003
Enteric viruses	<i>Legionella spp., Mycobacterium spp., and Leptospira</i>	-	-	Toze, 2006
Enteric viruses	<i>Clostridium perfringens, Escherichia coli</i>	culture	conventional	Payment et al. 2001
	<i>Salmonella enterica, E. coli, V. cholerae, Yersinia pestis, M. tuberculosis, B. anthracis, Streptococcus agalactiae</i>	Illumina MiSeq	conventional	Kumaraswamy et al. 2014

Table 4.3 (cont'd)

	<i>M.tuberculosis, C. perfringens, E. faecalis, Legionella pneumophila, S. flexneri, S. pyogenes, V. cholerae, S. enterica, S. pneumoniae, S. dysenteriae, N. meningitides, and Y. pestis</i>	Illumina Hiseq 2000	conventional	Cai and Zhang, 2013
	<i>Aeromonas, Arcobacter, Clostridium, Corynebacterium, Legionella, Leptospira, Pseudomonas, Streptococcus</i>	PCR, pyrosequencing	MBR and conventional	Ye and Zhang, 2011
	<i>Mycobacterium fortuitum, Mycobacterium phlei, Mycobacterium chelonae, Clostridium perfringens</i>	pyrosequencing	conventional	Bibby et al. 2010
	<i>A. hydrophila, B. cereus, C. perfringens, E. faecalis, E. coli, K. pneumoniae, P. aeruginosa, Salmonella spp.</i>	DNA microarray, QPCR,	conventional	Lee et al. 2008
	<i>A. hydrophila, C.perfringens, Enterococcus faecalis, Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa</i>	QPCR	conventional	Shannon et al. 2007
	<i>A. hydrophila, B. cereus, C. perfringens, E. faecalis, E. coli O158:H7, K. pneumoniae, P. aeruginosa, Salmonella sp.</i>	DNA microarray, QPCR,	conventional	Lee et al. 2006

Table 4.4: Wastewater Treatment Plant Characteristics

	EAST LANSING WWTP	TRAVERSE CITY WWTP
Wastewater treatment process (Biological treatment)	Activated Sludge (AS)	Membrane Biological Reactor (MBR)
Sludge Retention Time (SRT)	14 days	7.58 days
Capacity	18.8 MGD*	17.0 MGD
Average flow	13.4 MGD	8.5 MGD
Discharge Rate	14.1 MGD	4.0 MGD
Disinfection	<i>Chlorine (Cl)</i>	<i>Ultra-Violet (UV)</i>

*MGD-Millions gallon per day

Table 4.5: Metagenome analysis statistics for virus samples (by MetaVir).

Sample No.	Metagenome Name	Number of contigs	Ratio of affiliated sequences	No. of genes predicted
East Lansing Waste Water Treatment Plant (EL)				
S2	EL_AD_VIRUS	256064	11.6	365870
S4	EL_BD_VIRUS	151994	19.39	258135
Traverse City Waste Water Treatment Plant (TC)				
S9	TC_AD_VIRUS	197517	16.36	309715
S11	TC_BD_VIRUS	151992	19.85	258140

Table 4.6a: Metagenome analysis statistics for bacterial samples (generated by MGRAST).

Sample No.	Metagenome Name	Raw bp count	No. of contigs	Mean Sequence Length (bp)	Post QC: bp Count	Post QC No. of contigs	Post QC: Mean Sequence Length (bp)
East Lansing Waste Water Treatment Plant (EL)							
S1	EL_AD	176,512,330	238657	739 ± 1927	147771487	234355	630 ± 482
S3	EL_BD	184,817,243	256881	719 ± 1374	153515277	251116	611 ± 411
Traverse City Waste Water Treatment Plant (TC)							
S8	TC_AD	206,030,297	288930	713 ± 2198	176537661	285235	618 ± 484
S10	TC_BD	179,323,360	259220	691 ± 1048	14,645886	252517	584 ± 339

Table 4.6b: Functional category Hit distribution (Bacterial samples).

Metagenome Name	Processed: Predicted Protein Features	Processed: Predicted rRNA Features	Alignment: Identified Protein Features	Alignment: Identified rRNA Features
EL_AD	270,966	17,827	172,766	758
EL_BD	297,712	22,251	167,031	625
TC_AD	341,222	23,529	218,046	546
TC_BD	297,052	19,752	191,448	499

Note: Sequences were assembled and contigs generated were analyzed on MGRAST; Abbreviation: bp= base pair; AD=After Disinfection Effluent; BD= Before Disinfection Effluent

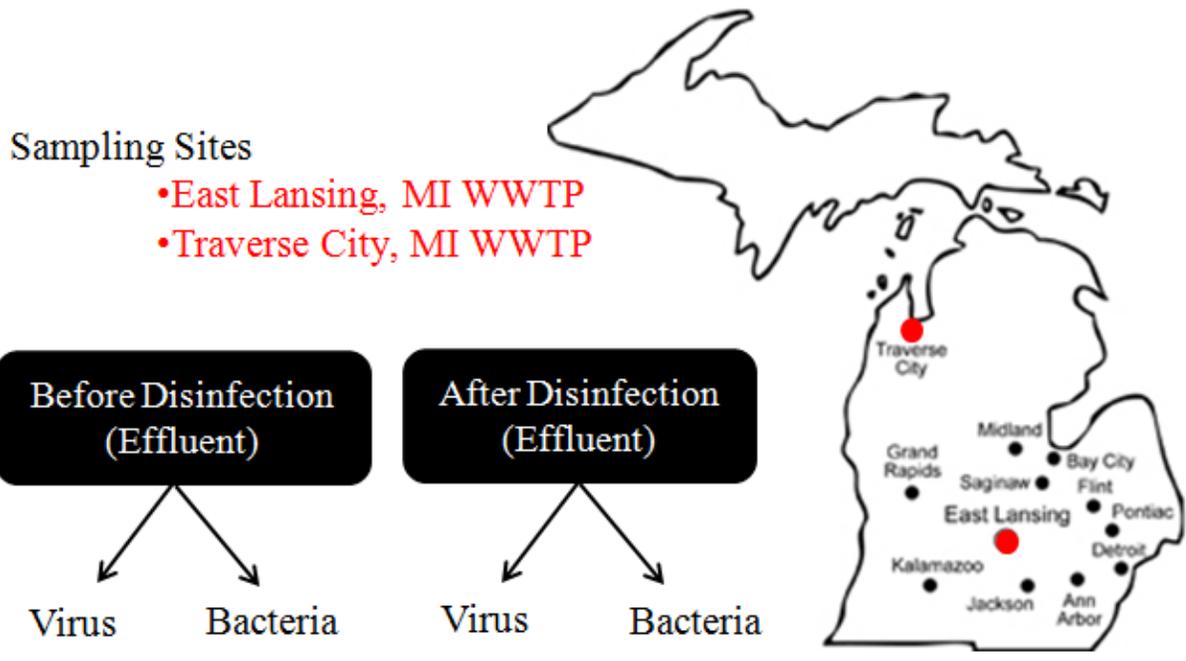


Figure 4.1: Location of two wastewater treatment plants selected for effluent sampling for investigation of the viral and bacterial community.

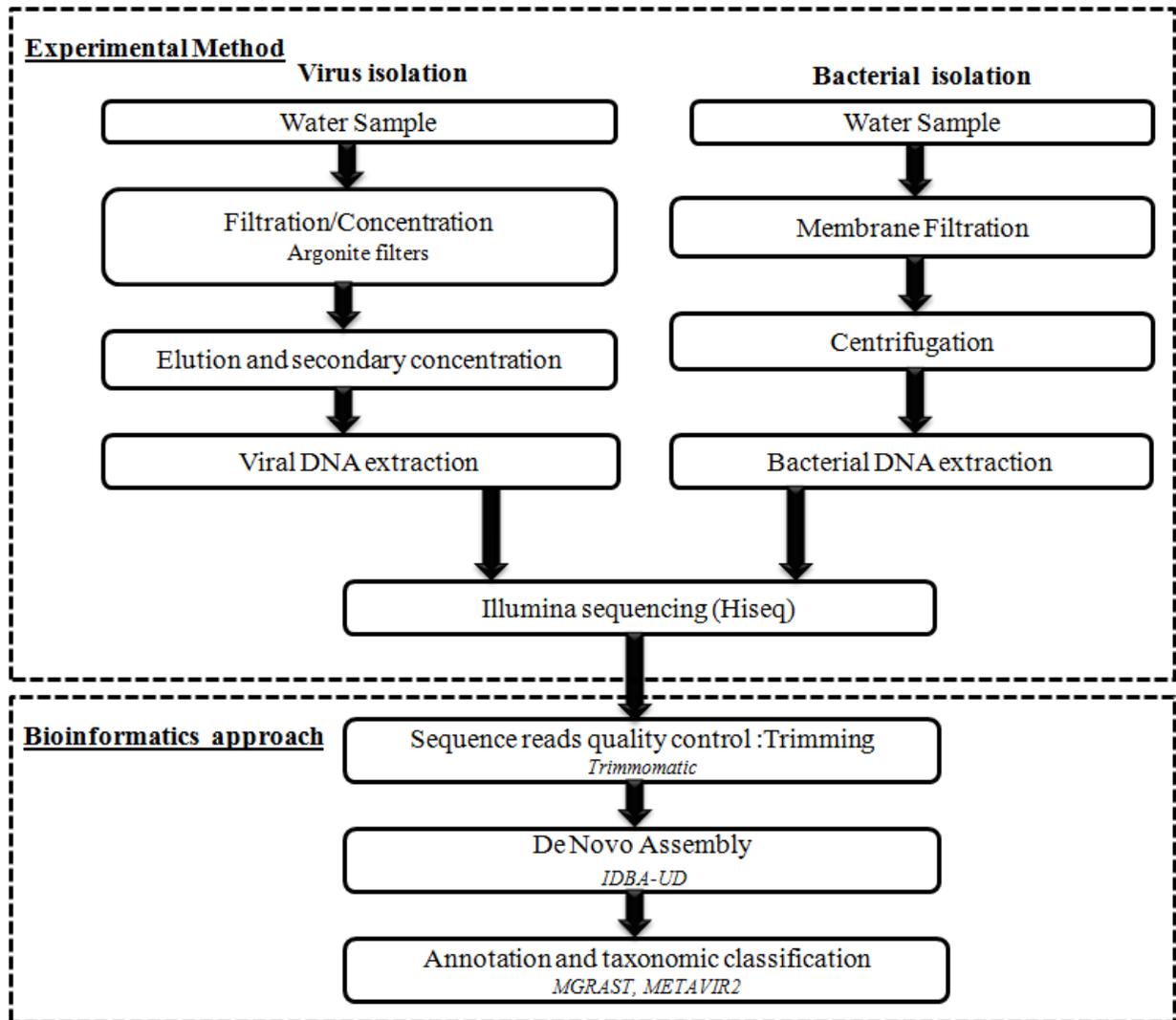
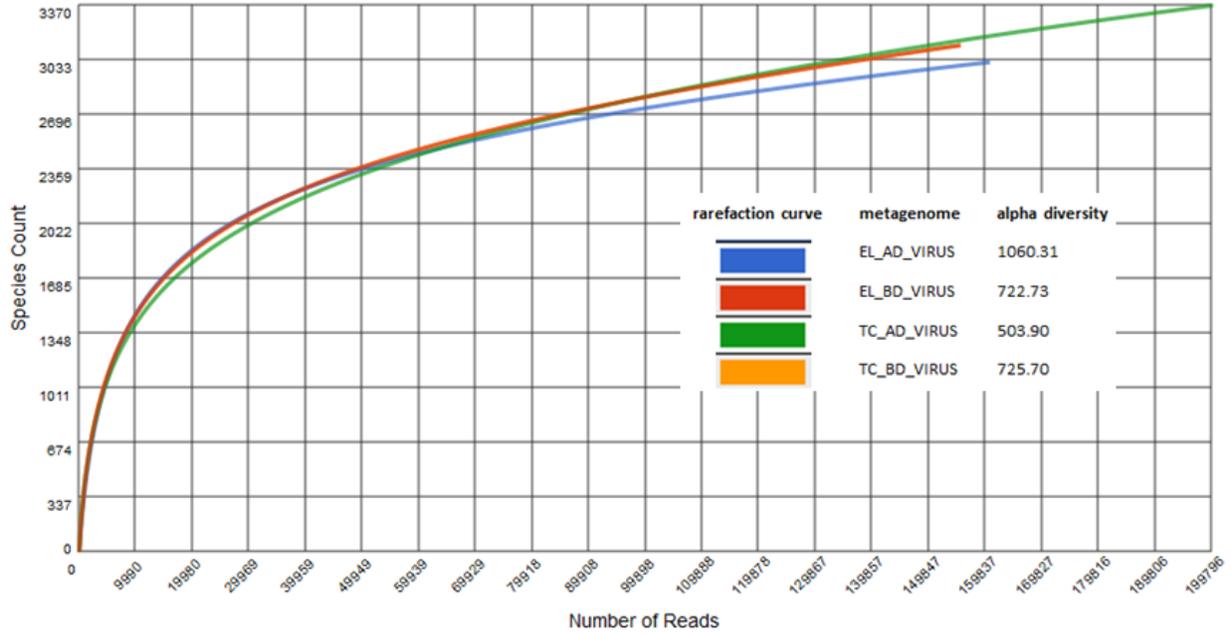
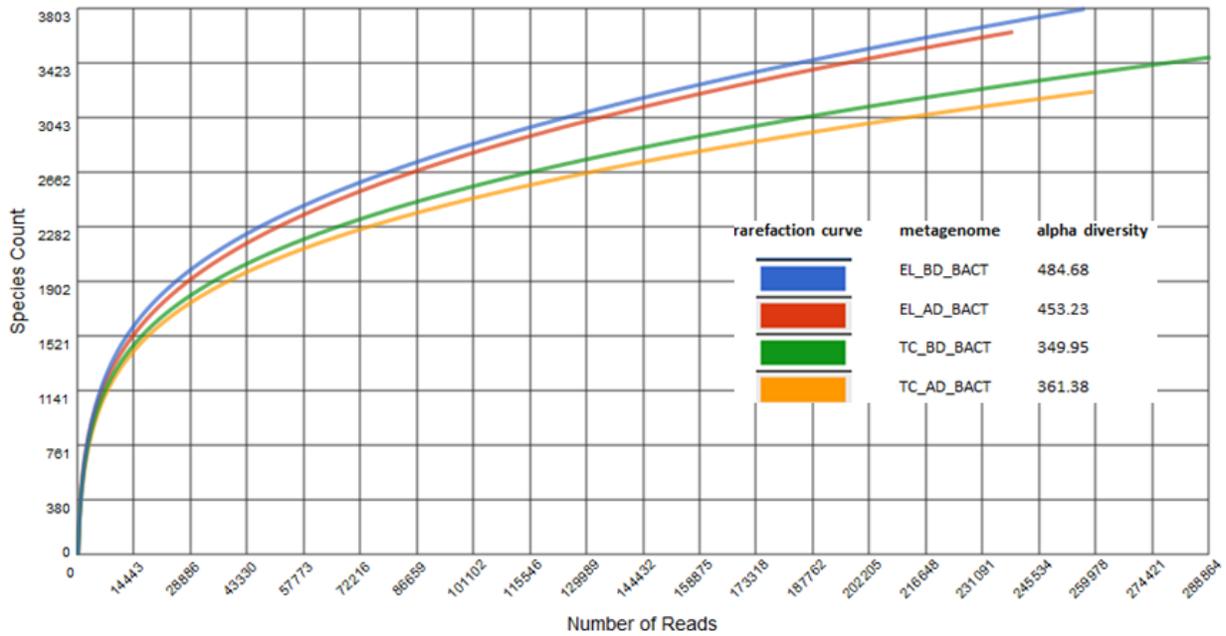


Figure 4.2: Schematic flowchart showing the procedure followed for metagenomic analysis



(a) Rarefaction curve for virus enriched samples in effluent samples



(b) Rarefaction curve for bacterial enriched samples in effluent samples

Figure 4.3: Rarefaction curve of species richness in virus and bacterial DNA enriched samples from effluents of two different WWTPs. Note: S2: EL_AD_VIRUS=East Lansing after disinfection effluent Virus sample, S4: EL_BD_VIRUS= East Lansing before disinfection effluent Virus sample; S9: TC_AD_VIRUS= East Lansing after disinfection effluent Virus sample; S11: TC_BD_VIRUS= East Lansing before disinfection effluent Virus sample

Table 4.7: Taxonomic comparison heat map based contigs best BLAST hit ratios (number of hits for the genome divided by total number of hits in the metagenome).

Taxonomy	S2 EL_AD	S4 EL_BD	S9 TC_AD	S11 TC_BD
Retro-transcribing viruses	0.27	0.08	0.08	0.07
Caulimoviridae	0.06	0.03	0.04	0.03
Retroviridae	0.21	0.05	0.05	0.04
dsDNA viruses, no RNA stage	96.39	97.45	97.01	96.71
Adenoviridae	0.01	0.02	0	0.01
Ascoviridae	0.83	0.15	0.29	0.14
Asfarviridae	0.06	0.13	0.2	0.09
Baculoviridae	0.56	0.2	0.31	0.17
Bicaudaviridae	0.12	0.05	0.04	0.04
Caudovirales	61.2	78.11	75.05	78.08
Corticoviridae	0.01	0.01	0.01	0.01
Herpesvirales	0.4	0.21	0.14	0.22
Iridoviridae	0.7	0.44	0.24	0.49
Ligamenvirales	0.08	0	0	0.04
Lipothrixviridae	0.07	0.02	0.03	0.02
Marseilleviridae	0.8	0	0	0.34
Mimiviridae	10.26	2.76	3.3	3.75
Nimaviridae	0.01	0.03	0.01	0.03
Nudiviridae	0.12	0	0	0.06
Phycodnaviridae	12.26	4.85	5.97	4.62
Plasmaviridae	0.01	0	0.01	0
Polydnaviridae	0.02	0.01	0.01	0.01
Poxviridae	1.57	0.54	0.77	0.49
Rudiviridae	0.01	0.01	0.01	0.01
Tectiviridae	0.04	0.03	0.08	0.04
unclassified dsDNA phages	2.76	5.82	5.81	5.52
unclassified dsDNA viruses	4.57	4.06	4.71	2.56
dsRNA viruses	0.01	0.01	0.01	0.01
Cystoviridae	0.01	0.01	0.01	0.01
ssDNA viruses	0.55	0.15	0.46	0.14
Circoviridae	0	0	0.01	0
Inoviridae	0.52	0.12	0.22	0.11

Table 4.7 (cont'd)

Microviridae	0.03	0.02	0.03	0.02
Parvoviridae	0	0	0.15	0
unclassified ssDNA viruses	0.01	0.01	0.05	0.01
ssRNA viruses	0.07	0.03	0.03	0.03
ssRNA positive-strand viruses, no DNA stage	0.07	0.03	0.03	0.03
unassigned viruses	0.05	0.02	0.02	0.02
unclassified archaeal viruses	0.02	0.01	0	0.01
unclassified phages	2.22	2.22	2.36	2.74
unclassified virophages	0.01	0.02	0.03	0.02
unclassified viruses	0.42	0	0	0.25

Note: Comparison of the taxonomic compositions computed from a BLAST comparison with ncbi refseq complete viral genomes proteins using BLASTp (threshold of 50 on the BLAST bitscore). Abbreviation: S2: EL_AD =East Lansing after disinfection effluent virus sample, S4: EL_BD = East Lansing before disinfection effluent virus sample; S9: TC_AD = East Lansing after disinfection effluent virus sample; S11: TC_BD = East Lansing before disinfection effluent virus sample

Table 4.8: Organism Abundance (Bacteria Phylum Distribution).

Sample Type	EL_AD	EL_BD	TC_AD	TC_BD
Proteobacteria	0.989	0.973	1	0.983
acteroidetes	0.777	0.735	0.802	0.811
Firmicutes	0.571	0.591	0.605	0.582
Actinobacteria	0.541	0.623	0.654	0.62
Cyanobacteria	0.517	0.539	0.564	0.541
Planctomycetes	0.515	0.644	0.617	0.617
Verrucomicrobia	0.513	0.542	0.659	0.65
Nitrospirae	0.492	0.345	0.582	0.59
Acidobacteria	0.433	0.482	0.517	0.495
Chloroflexi	0.43	0.509	0.537	0.495
Chlorobi	0.406	0.436	0.471	0.456
unclassified (derived from Bacteria)	0.386	0.411	0.428	0.408
Deinococcus-Thermus	0.34	0.39	0.416	0.386
Chlamydiae	0.34	0.371	0.251	0.225
Gemmatimonadetes	0.316	0.449	0.358	0.303
Spirochaetes	0.286	0.362	0.444	0.384
Aquificae	0.247	0.253	0.306	0.279
Poribacteria	0.236	0.085	0.317	0.327
Fusobacteria	0.229	0.234	0.221	0.217
Synergistetes	0.191	0.204	0.219	0.17
Deferribacteres	0.175	0.182	0.195	0.175
Thermotogae	0.155	0.228	0.264	0.231

Table 4.8 (cont'd)

Lentisphaerae	0.127	0.205	0.245	0.243
Chrysiogenetes	0.119	0.136	0.131	0.12
Dictyoglomi	0.087	0.087	0.104	0.104
Elusimicrobia	0.057	0.029	0.087	0.062
Tenericutes	0.042	0.101	0.08	0.069
Fibrobacteres	0	0.05	0.039	0.05

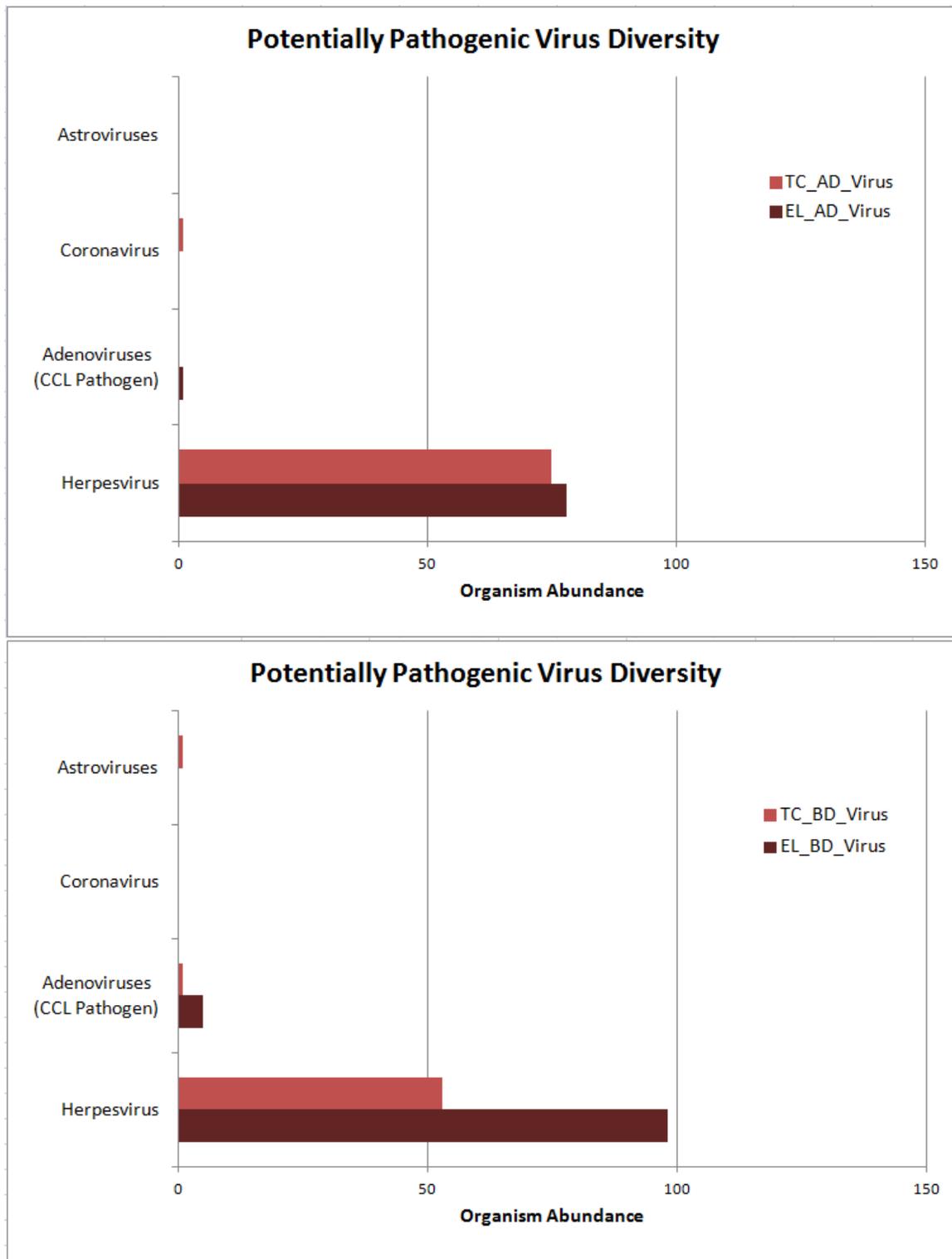


Figure 4.4: Potentially pathogenic virus abundance in effluent samples. Adenoviruses are the only CCL pathogens found in our samples. Note: The scale has been log transformed for better visual.

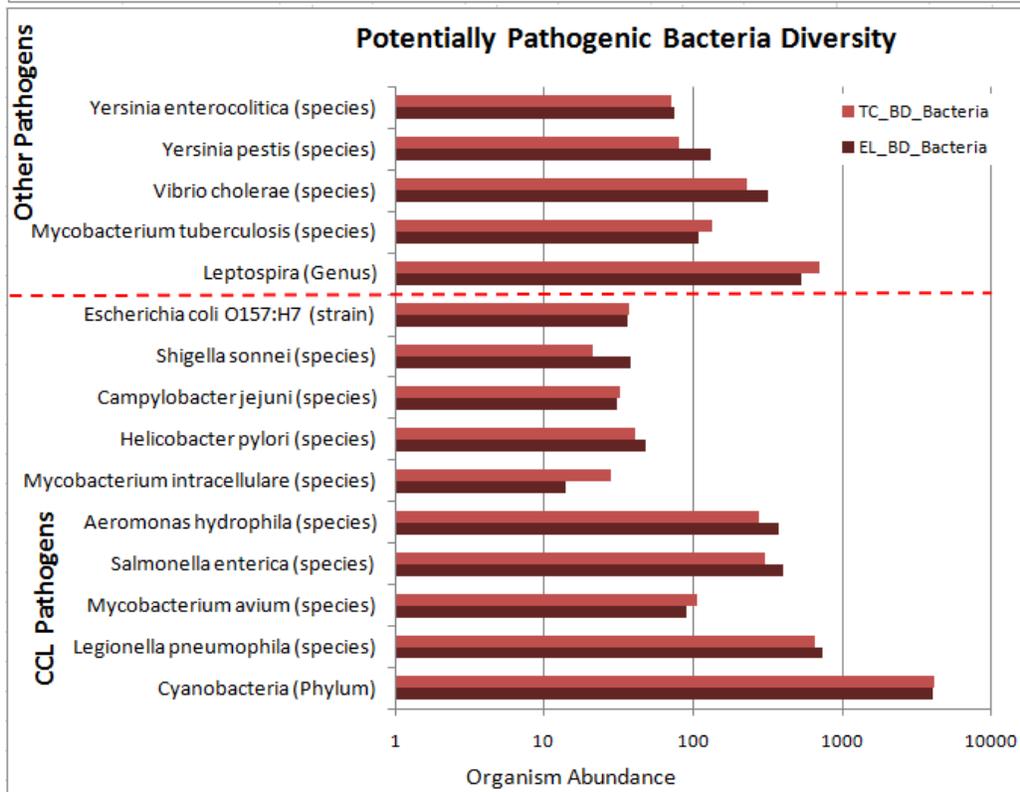
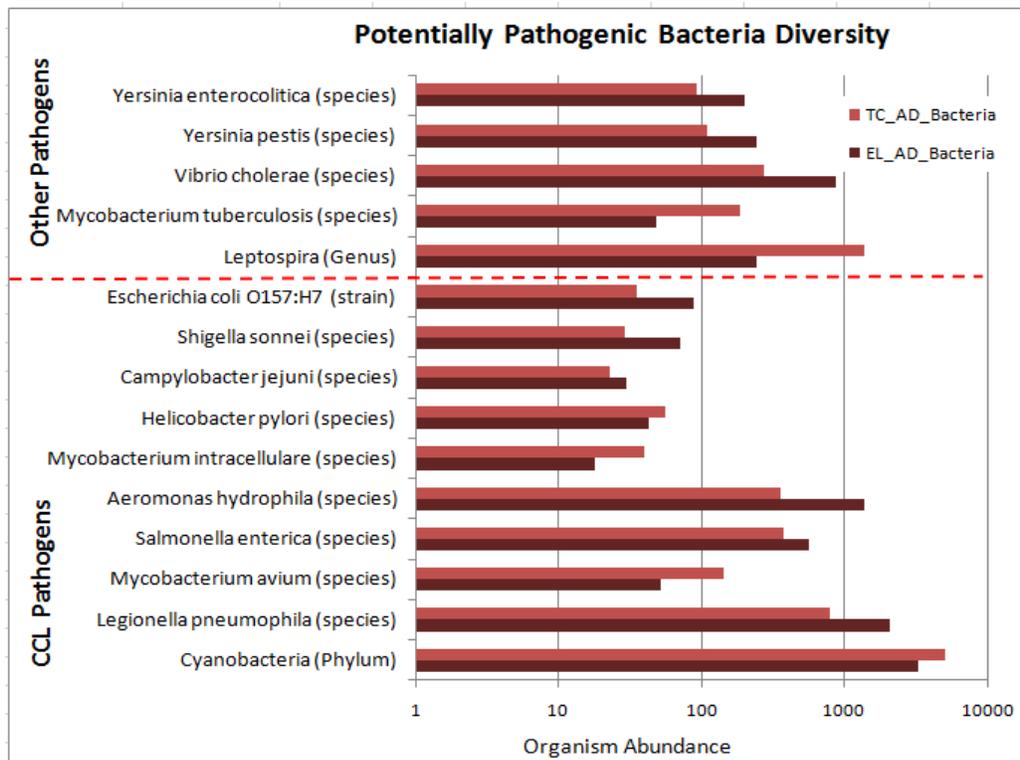


Figure 4.5: Potentially pathogenic bacteria diversity abundance in effluent samples. Note: The scale has been log transformed for better visual.

REFERENCES

REFERENCES

1. Albinana-Gimenez, N., Miagostovich, M. P., Calgua, B., Huguet, J. M., Matia, L., & Girones, R. (2009). Analysis of adenoviruses and polyomaviruses quantified by qPCR as indicators of water quality in source and drinking-water treatment plants. *Water research*, 43(7), 2011-2019.
2. Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R., & Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *Journal of microbiological methods*, 92(2), 183-188.
3. Aw, T. G., Howe, A., & Rose, J. B. (2014). Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *Journal of virological methods*.
4. Bibby, K. (2013). Metagenomic identification of viral pathogens. *Trends in biotechnology*, 31(5), 275-279.
5. Bibby, K., & Peccia, J. (2013a). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental science & technology*, 47(4), 1945-1951.
6. Bibby, K., & Peccia, J. (2013b). Prevalence of respiratory adenovirus species B and C in sewage sludge. *Environ. Sci.: Processes Impacts*, 15(2), 336-338.
7. Bibby, K., Viau, E., & Peccia, J. (2010). Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. *Water research*, 44(14), 4252-4260.
8. Bibby, K., Viau, E., & Peccia, J. (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Letters in applied microbiology*, 52(4), 386-392.
9. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170
10. Cai, L., & Zhang, T. (2013). Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environmental science & technology*, 47(10), 5433-5441.
11. Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., ... & Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *MBio*, 2(5), e00180-11.
12. Costán-Longares, A., Montemayor, M., Payán, A., Méndez, J., Jofre, J., Mujeriego, R., & Lucena, F. (2008). Microbial indicators and pathogens: removal, relationships and predictive capabilities in water reclamation facilities. *Water research*, 42(17), 4439-4448.

13. da Silva, A. K., Le Saux, J. C., Parnaudeau, S., Pommepuy, M., Elimelech, M., & Le Guyader, F. S. (2007). Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: different behaviors of genogroups I and II. *Applied and Environmental Microbiology*, 73(24), 7891-7897.
14. Finkbeiner, S. R., Allred, A. F., Tarr, P. I., Klein, E. J., Kirkwood, C. D., & Wang, D. (2008). Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS pathogens*, 4(2), e1000011.
15. Fong, T. T., Phanikumar, M. S., Xagorarakis, I., & Rose, J. B. (2010). Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a Michigan river. *Applied and environmental microbiology*, 76(3), 715-723.
16. Francy, D. S., Stelzer, E. A., Bushon, R. N., Brady, A. M., Williston, A. G., Riddell, K. R., ... & Gellner, T. M. (2012). Comparative effectiveness of membrane bioreactors, conventional secondary treatment, and chlorine and UV disinfection to remove microorganisms from municipal wastewaters. *Water research*, 46(13), 4164-4178.
17. Francy, D.S., Stelzer, E.A., Bushon, R.N., Brady, A.M.G., Mailot, B.E., Spencer, S.K., Borchardt, M.A., Elber, A.G., Riddell, K.R., and Gellner, T.M. (2011). Quantifying viruses and bacteria in wastewater—Results, interpretation methods, and quality control: U.S. Geological Survey Scientific Investigations Report 2011–5150, 44 .
18. Gomez-Alvarez, V., Revetta, R. P., & Santo Domingo, J. W. (2012). Metagenomic analyses of drinking water receiving different disinfection treatments. *Applied and environmental microbiology*, 78(17), 6095-6102.
19. Griffin, D. W., Donaldson, K. A., Paul, J. H., & Rose, J. B. (2003). Pathogenic human viruses in coastal waters. *Clinical microbiology reviews*, 16(1), 129-143.
20. Health Canada (2013). Guidance on waterborne bacterial pathogens. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario (Catalogue No. H129-25/1-2014E-PDF). <http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/pathogens-pathogenes/index-eng.php#b.3.1>
21. Hewitt, J., Leonard, M., Greening, G. E., & Lewis, G. D. (2011). Influence of wastewater treatment process and the population size on human virus profiles in wastewater. *Water research*, 45(18), 6267-6276.
22. Hirani, Z. M., Bukhari, Z., Oppenheimer, J., Jjemba, P., LeChevallier, M. W., & Jacangelo, J. G. (2013). Characterization of effluent water qualities from satellite membrane bioreactor facilities. *Water research*, 47(14), 5065-5075.
23. Hu, M., Wang, X., Wen, X., & Xia, Y. (2012). Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresource technology*, 117, 72-79.

24. Hunt R. (2011). Herpes Viruses, Virology chapter eleven, University of south Carolina School of Medicine.
25. Jacangelo, J., Loughran, P., Petrik, B., Simpson, D., & McIlroy, C. (2003). Removal of enteric viruses and selected microbial indicators by UV irradiation of secondary effluent. *Water Science & Technology*, 47(9), 193-198.
26. Katayama, H., Haramoto, E., Oguma, K., Yamashita, H., Tajima, A., Nakajima, H., & Ohgaki, S. (2008). One-year monthly quantitative survey of noroviruses, enteroviruses, and adenoviruses in wastewater collected from six plants in Japan. *Water research*, 42(6), 1441-1448.
27. Kitajima, M., Iker, B. C., Pepper, I. L., and Gerba, C. P. (2014). Relative abundance and treatment reduction of viruses during wastewater treatment processes—identification of potential viral indicators. *Science of the Total Environment*, 488, 290-296. Kowal, N. E. (1985). *Health effects of land application of municipal*. EPA/1-85/015 ss.
28. Kumaraswamy, R., Amha, Y. M., Anwar, M. Z., Henschel, A., Rodríguez, J., & Ahmad, F. (2014). Molecular analysis for screening human bacterial pathogens in municipal wastewater treatment and reuse. *Environmental science & technology*, 48(19), 11610-11619.
29. Kuo, D. H. W., Simmons, F. J., Blair, S., Hart, E., Rose, J. B., & Xagorarakis, I. (2010). Assessment of human adenovirus removal in a full-scale membrane bioreactor treating municipal wastewater. *Water research*, 44(5), 1520-1530.
30. La Rosa, G., Pourshaban, M., Iaconelli, M., & Muscillo, M. (2010). Quantitative real-time PCR of enteric viruses in influent and effluent samples from wastewater treatment plants in Italy. *Annali dell'Istituto superiore di sanità*, 46(3), 266-273.
31. La Rosa, G., Pourshaban, M., Iaconelli, M., & Muscillo, M. (2010). Quantitative real-time PCR of enteric viruses in influent and effluent samples from wastewater treatment plants in Italy. *Annali dell'Istituto superiore di sanità*, 46(3), 266-273.
32. Lee, D. Y., Lauder, H., Cruwys, H., Falletta, P., & Beaudette, L. A. (2008). Development and application of an oligonucleotide microarray and real-time quantitative PCR for detection of wastewater bacterial pathogens. *Science of the total environment*, 398(1), 203-211.
33. Lee, D. Y., Shannon, K., & Beaudette, L. A. (2006). Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. *Journal of microbiological methods*, 65(3), 453-467.
34. Malik, Y. S., & Matthijnsens, J. (2014). Enteric viral infection in human and animal. *VirusDisease*, 25(2), 145-146.

35. McLellan, S. L., Huse, S. M., Mueller-Spitz, S. R., Andreishcheva, E. N., and Sogin, M. L. (2010). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental microbiology*, 12(2), 378-392.
36. Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R. and Wilke, A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* 9(1), 386.
37. Miller, R. R., Montoya, V., Gardy, J. L., Patrick, D. M., & Tang, P. (2013). Metagenomics for pathogen detection in public health. *Genome medicine*, 5(9), 81.
38. Naccache, S. N., Federman, S., Veeeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., ... & Chiu, C. Y. (2014). A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Research*.
39. Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, et al. (2009) Direct Metagenomic Detection of Viral Pathogens in Nasal and Fecal Specimens Using an Unbiased High-Throughput Sequencing Approach. *PLoS ONE* 4(1): e4219.
40. Odjadjare, E. E., Obi, L. C., & Okoh, A. I. (2010). Municipal wastewater effluents as a source of listerial pathogens in the aquatic milieu of the eastern cape province of South Africa: a concern of public health importance. *International journal of environmental research and public health*, 7(5), 2376-2394.
41. Okoh, A. I., Odjadjare, E. E., Igbinsosa, E. O., & Osode, A. N. (2007). Wastewater treatment plants as a source of microbial pathogens in receiving watersheds. *African Journal of Biotechnology*, 6(25).
42. Payment, P., Plante, R., & Cejka, P. (2001). Removal of indicator bacteria, human enteric viruses, Giardia cysts, and Cryptosporidium oocysts at a large wastewater primary treatment facility. *Canadian Journal of Microbiology*, 47(3), 188-193.
43. Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420-1428.
44. Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., Daroub S. H. , Camargo F. AO., Farmerie W. G. and Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, 1(4), 283-290. Sewage
45. Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., and Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental microbiology*, 11(11), 2806-2820.

46. Roux, S., Tournayre, J., Mahul, A., Debroas, D., & Enault, F. (2014). Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC bioinformatics*, 15(1), 76.
47. Shannon, K. E., Lee, D. Y., Trevors, J. T., & Beaudette, L. A. (2007). Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Science of the total environment*, 382(1), 121-129.
48. Sidhu, J., Hanna, J., & Toze, S. (2008). Survival of enteric microorganisms on grass surfaces irrigated with treated effluent. *Journal of water and health*, 6(2), 255-262.
49. Sima, L. C., Schaeffer, J., Le Saux, J. C., Parnaudeau, S., Elimelech, M., & Le Guyader, F. S. (2011). Calicivirus removal in a membrane bioreactor wastewater treatment plant. *Applied and environmental microbiology*, 77(15), 5170-5177.
50. Simmons, F. J., & Xagorarakis, I. (2011). Release of infectious human enteric viruses by full-scale wastewater utilities. *Water research*, 45(12), 3590-3598.
51. Simmons, F. J., & Xagorarakis, I. (2011). Release of infectious human enteric viruses by full-scale wastewater utilities. *Water research*, 45(12), 3590-3598.
52. Simmons, F. J., Kuo, D. H. W., & Xagorarakis, I. (2011). Removal of human enteric viruses by a full-scale membrane bioreactor during municipal wastewater processing. *Water research*, 45(9), 2739-2750.
53. Simmons, F. J., Kuo, D. H. W., & Xagorarakis, I. (2011). Removal of human enteric viruses by a full-scale membrane bioreactor during municipal wastewater processing. *Water research*, 45(9), 2739-2750.
54. Snaidr, J., Amann, R., Huber, I., Ludwig, W., and Schleifer, K. H. (1997). Phylogenetic analysis and in situ identification of bacteria in activated sludge. *Applied and Environmental Microbiology*, 63(7), 2884-2896.
55. Tamaki, H., Zhang, R., Angly, F. E., Nakamura, S., Hong, P. Y., Yasunaga, T., ... & Liu, W. T. (2012). Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environmental microbiology*, 14(2), 441-452.
56. Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. *Microb Inform Exp*, 2(3).
57. Toze, S. (1997). Microbial Pathogens in Wastewater: Literature review for urban water systems multi-divisional research program. CSIRO Land and Water.
58. Toze, S. (1999). PCR and the detection of microbial pathogens in water and wastewater. *Water Research*, 33(17), 3545-3556.
59. Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature reviews genetics*, 6(11), 805-814.

60. USEPA, 2001. Manual of Methods for Virology (Chapter 14). EPA 600/4–84/013 Office of Water, U.S. Environmental Protection Agency, Washington, DC.
61. Wilke, A., Glass, E. M., Bischof, J., Braithwaite, D., DSouza, M., Gerlach, W., ... & Meyer, F. (2013). MG-RAST Technical report and manual for version 3.3. 6–rev.
62. Wommack, K. E., Srinivasiah, S., Liles, M., Bhavsar, J., Bench, S., Williamson, K. E., & Polson, S. W. (2011). Metagenomic contrasts of viruses in soil and aquatic environments. *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*. New York: John E. Wiley & Sons.
63. Xia, S., Duan, L., Song, Y., Li, J., Piceno, Y. M., Andersen, G. L., Alvarez-Cohen L., Moreno-Andrade I., Huang C., and Hermanowicz, S. W. (2010). Bacterial community structure in geographically distributed biological wastewater treatment reactors. *Environmental science & technology*, 44(19), 7391-7396.
64. Yang, J., Yang, F., Ren, L., Xiong, Z., Wu, Z., Dong, J., ... & Jin, Q. (2011). Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *Journal of clinical microbiology*, 49(10), 3463-3469.
65. Yates, V. (2011). Pathogens in reclaimed water. University of California. Riverside (<http://www.geoflow.com/wastewater/pathogens.htm>)
66. Ye, L., & Zhang, T. (2011). Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. *Environmental science & technology*, 45(17), 7173-7179.
67. Ye, L., & Zhang, T. (2011). Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. *Environmental science & technology*, 45(17), 7173-7179.
68. Zhang, C. M., & Wang, X. C. (2014). Distribution of Enteric Pathogens in Wastewater Secondary Effluent and Safety Analysis for Urban Water Reuse. *Human and Ecological Risk Assessment: An International Journal*, 20(3), 797-806.
69. 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), e26041.
70. Zhou, J., Wang, X. C., Ji, Z., Xu, L., & Yu, Z. (2014). Source identification of bacterial and viral pathogens and their survival/fading in the process of wastewater treatment, reclamation, and environmental reuse. *World Journal of Microbiology and Biotechnology*, 1-12.

CHAPTER 5

VIRUS AND PHAGE DIVERSITY IN ACTIVATED SLUDGE FROM A CONVENTIONAL AND A MBR WASTEWATER UTILITY USING METAGENOMIC ANALYSES

Mariya Munir, Camille McCall, Terence Marsh, and Irene Xagorarakis. (in preparation)

Abstract

In this paper, phage and virus diversity of two different wastewater treatment systems, conventional and membrane bioreactor (MBR) utilities were investigated through metagenomics. Samples collected from activated sludge were studied using Illumina Hiseq sequencing. Analysis software MGRAST v3.0 pipeline was used for analyzing the assembled sequences.

Phycodnaviridae, Herpesviridae, Mimiviridae and Lipothrixviridae were present in high numbers in sludge samples. Most of the sequences detected were uncharacterized; indicating a greater number of viral diversity is yet to be discovered.

Siphoviridae, Podoviridae and Myoviridae are the most common group of phages found in both the phage enriched samples from sludge obtained from the two treatment plants. On genus and species level, our results showed differences in viral community compositions between two different WWTPs.

Keywords: Metagenomics, activated sludge, virus, phage, diversity, Illumina Hiseq, MBR

1. Introduction

Viruses are small infectious particles consisting of a nucleic acid core (single or double stranded RNA or DNA) typically 20–200 nm in size, enclosed by a protein coat (capsid) and in some cases a lipid envelope (Carter and Saunders, 2007). Viruses are potentially the most

important and most hazardous among the pathogens found in wastewater (Sidhu et al., 2008, Toze, 1997). They are also generally more difficult to detect in environmental samples. The huge diversity of viruses that exist in human populations are potentially excreted and concentrated in wastewater (Bibby and Peccia, 2013). It has been shown that viruses persist in the water environment for extended periods and are also more resistant to removal by wastewater treatment systems compared to bacteria (Aw et al. 2014, Gomila et al., 2008). According to Xagorarakis et al (2014), presence of viruses in water and wastewater is a complex problem for environmental engineers because of prevalence, infectivity, and resistance of viruses to disinfection.

Bacteriophages are the most abundant biological entities on the planet (Bergh et al., 1989) and are regarded as an active part of the activated sludge microbial ecosystem (Khan et al. 2002). Bacteriophages (phages) are the viruses that infect bacteria. They are the most dominating and abundant among viral diversity (Bibby, 2014). They are associated with the large bacterial (host) populations. Like all viruses, phages are obligate intracellular parasites and require the metabolic machinery of the host cell to support their reproduction (Withey et al. 2005). The role of bacteriophages in the emergence of novel bacterial pathogens by horizontal gene transfer is noteworthy. Very little information is available regarding phage-mediated transduction (Colomer-Lluch et al. 2011, Sander and Schmieger 2001). The significance of bacteriophages for the process of transduction among bacterial population has been undermined for long. Concentration of 10^8 to 10^9 phage-like-particles per ml was reported in different full scale activated sludge bioreactors (Ostawa et al. 2007).

It is suggested that phage abundance in activated sludge is higher than any other environment (Ostawa et al. 2007, Shapiro and Kushmaro 2011, Rosenberg et al. 2010, Wu and

Liu 2009). Occurrences of different viable bacteriophages in the wastewater environment (somatic coliphage range: 10^3 ml^{-1} to 10^5 ml^{-1}) have been reported to be typically lower (Muniesa et al. 2004b, Mandilara et al. 2006, Khan et al. 2002, Havelaar et al. 1990). An increase in total phage concentration was observed during an activated sludge process, suggesting active replication was occurring via host infection and lysis (Ewert and Paynter 1980). It was shown that in WWTP the phage to bacterial cell ratio was approximately 10:1 (Rosenberg et al. 2010).

The objective of this study was to describe and compare the diversity viruses and phages in a conventional and MBR (membrane biological Reactor) utility using metagenomics with high throughput sequencing technology (Hiseq Illumina sequencing). Not many studies have provided complete diversity in wastewater using metagenomics based on our literature review. This study provides a complete diversity characterization of viruses and phage- enriched DNA in activated sludge from two different wastewater treatment plants.

2. Materials and Methods

2.1. Sample Collection: Effluent samples before and after disinfection and activated sludge were collected from East Lansing (EL) WWTP and Traverse City (TC) WWTP in Michigan (U.S.A.) in 2013. The characteristic of these WWTPs is shown in Table 5.1. Activated sludge (AS) sample was collected in two 1L nalgene bottles, mixed together in laboratory and then separated into sections for phage and virus isolation from each WWTP. Samples were kept on ice and were transported to the Water Quality Engineering Laboratory at Michigan State University (East Lansing, U.S.A.) for further immediate processing. The schematic of the all the methods used in this study is presented in Figure 5.1.

2.2. Sample Processing:

2.2.1. Virus elution/isolation process:

The virus elution and concentration from sludge samples were performed according to the ASTM-4994 using 10% beef extract (USEPA, 2001). Briefly, sludge samples were conditioned with 0.05M AlCl₃ and the pH adjusted to 3.5 ± 0.1 using 1 M HCl with 30 min mixing, to ensure all viruses adsorbed to the solids, before elution process. Conditioned samples were then poured into centrifuge bottles and centrifuge for 15 min at 2500×g at 4 °C. To each decanted bottle, 10% of beef extract was added and mixed for 30 min. This was followed by another centrifuge at 10000×g at 4 °C for 30 min. This procedure was followed by organic flocculation where the pH of the eluted beef extract is brought down to 3.5 ± 0.1 with constant slow stirring for 30 min. Another round of centrifuge at 2500×g for 15 min at 4 °C was done, discarding the supernatant and the pellets were resuspended and dissolved in 0.01M phosphate buffer saline (PBS). The re-suspension was transferred to a small sterile beaker with stir bar for vigorously stirring to dissolve the flocculate. Antibiotics (Kanamycin, Gentamicin or Antimucotic: 1mL) is added at this step and pH is adjusted to 7.0-7.5 with 1N HCL or 1N NaOH. One last optional centrifuge, to remove debris for easy filtering, was done for 10 min at 1850×g in 50 mL centrifuge tubes. As before, the supernatant was loaded into a 60 mL syringe and passed through a 0.22 μm sterilized filter for removal of bacteria, fungi and other contaminating agents. All samples were completely mixed and placed into 2 mL cryogenic tubes and stored at -80 °C until further analysis.

2.2.2. Phage isolation:

Phage isolation from activated sludge samples was carried out using previously developed method in our lab. Mitomycin C (1μg/ml) was added to induce the samples and the samples were incubated at room temperature for 24 hrs while gently shaking (150 rpm). Several drops of

chloroform were added to the samples to complete lysis and incubation was continued for another 15-30 mins. The samples (250-300mL of sludge) were then centrifuged at 3396 xg for 45 minutes in F10S-6X500Y rotor and the supernatant was carefully decanted. Each supernatant was filtered through a 0.22 µm filter (Millipore, Billerica, MA) and then the bacteriophages were precipitated with PEG-NaCl (Colomer-Lluch et al. 2011, Sander and Schmieger 2001, Muniesa et al. 2004a). The PEG precipitate was collected by centrifugation at 10000xg in an aerosol-tight fixed-angle rotor at 4°C for 40 minutes. The supernatant was carefully decanted and the pellet was resuspended in 1.0 ml of SM buffer (Yamamota et al., 1970). Any free DNA that co-purified with bacteriophage was removed by digestion with DNase I (100 Units/mL) (Colomer-Lluch et al. 2011). The phage preparations were stored at -80°C until DNA extraction was performed for molecular analysis.

2.3. Nucleic acid Extraction: Virus DNA and Phage DNA was extracted (from samples before and after disinfection) using a MagNA Pure Compact DNA extractor (Roche Applied Science, Indianapolis, IN, USA) following the protocol in the manufacturer's manual. The MagNA Pure Compact utilizes a magnetic-bead technology for the isolation process. Sample amount of 400 µL was loaded in the system and the elution volume was 100µL. The extracts were stored in a freezer at -20°C. Following extraction the quantity of bacterial and viral nucleic acid extracts from all samples were checked using the NanoDrop Spectrophotometer (NanoDrop® ND-1000, Wilmington, DE).

2.4. Metagenomic sequencing and analyses: Isolated DNA samples including phage-enriched DNA and virus-enriched DNA isolated from the sludge was sequenced on an Illumina platform (Illumina HiSeq, Roche Technologies) at The Research Technology Support Facility (RTSF) genomic center at Michigan State University generating 250 bp paired-end reads. Approximately

1 µg DNA (per core sample) was sent to the sequencing facility. The sequencing results were returned as .FASTQ.GZ files and they were converted to FASTQ files by processing them in MSU High Performance Computing Center (HPCC) secure shell (SSH) connection. The SSH connection was established by a widely used software known as PuTTY. In order to quality filter the illumina data, a flexible read trimming tool for Illumina NGS data called Trimmomatic was used for trimming Illumina data and removing adapters (Bolger et al. 2014). Finally, the trimmed sequences were assembled using an iterative De Bruijn Graph De Novo Assembler for Short Reads Sequencing data with Highly Uneven Sequencing Depth called IDBA-UD (Peng et al. 2012).

2.5. MG-RAST analyses: MetaGenome Rapid Annotation Subsystems Technology server (MG-RAST 3.3.1) was used to analyze the assembled data from each sample (Meyer et al. 2008). Each of the assembled data files were underwent quality control (QC) process. This included quality filtering, which involves removing sequences with ≥ 5 ambiguous base pairs, length filtering, removing sequences with length ≥ 2 standard deviations from the mean, and de-replication, which involves removing similar sequences that are artifacts of sequencing. The analysis obtained from MG-RAST consists of the phylogenetic comparison and functional annotation compared with the database. The results are expressed in the form of abundance profiles. The abundance represents an estimate of a number of sequences that contain a given annotation. This is found by multiplying each selected database match, or hit, by the number of representatives in each cluster. Hit refers to the number of unique database sequences that were found in the similarity search. The hit count can be smaller than the number of reads because of clustering, or larger than the number of reads due to double counting (Wilke et al. 2014). Data was analyzed based on organism abundance and on the functional distribution at the subsystem hierarchy with

maximum E- value cutoff of $1E^{-5}$, minimum percent identity cutoff of 60% and minimum alignment length cutoff of 15bps. The MG-RAST pipeline analysis includes the comparisons and functional annotations against the database. Each of the categories was further studied for detailed analysis and data was downloaded in excel sheet for further analysis.

3. Results and Discussion:

3.1. Metagenomic statistics:

Virus and phage diversity in wastewater treatment plants was analyzed using next generation sequencing (NGS) on the Illumina HiSeq platform. The study was conducted on four samples from two different WWTPs in Michigan. Virus and phage DNA was isolated from activated sludge (AS) from a conventional and MBR utility. Sequences generated were assembled using an IDBA-UD assembler. Metagenome analysis revealed that virus-enriched DNA sample contained 297613 and 244963 contigs sequences from ELWWTP and TCWWTP respectively, totaling 348 Mbp and phage-enriched DNA samples contained 154095 and 138306 ontigs sequences totaling 238 Mbp. Bioinformatic analysis statistics pre and post quality control process obtained from MGRAST server is presented in Table 5.2.

In virus isolated sample from EL, 7763 sequences (2.6%) failed to pass the QC pipeline. Of the sequences that passed QC, 363 sequences (0.1%) contain ribosomal RNA genes. Of the remainder, 185,201 sequences (62.2%) contain predicted proteins with known functions and 103,941 sequences (34.9%) contain predicted proteins with unknown function. 343 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. In TC virus isolated sample, 4211 sequences (1.7%) failed to pass the QC pipeline, 331 sequences (0.1%) contain ribosomal RNA genes, 152,188 sequences (62.1%) contain predicted proteins with known

functions and 87,863 sequences (35.9%) contain predicted proteins with unknown function. 368 (0.2%) of the sequences that passed QC have no rRNA genes or predicted proteins. Similarly, in phage isolated sample (EL), 16,775 sequences (10.9%) failed to pass the QC pipeline, 575 sequences (0.4%) contain ribosomal RNA genes, 57,544 sequences (37.3%) contain predicted proteins with known functions and 78,937 sequences (51.2%) contain predicted proteins with unknown function. 204 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. In TC phage isolated sample, 12,774 sequences (9.2%) failed to pass the QC pipeline, 364 sequences (0.3%) contain ribosomal RNA genes, 48,588 sequences (35.1%) contain predicted proteins with known functions and 76,405 sequences (55.2%) contain predicted proteins with unknown function. Only 144 (0.1%) of the sequences that passed QC have no rRNA genes or predicted proteins. Figure 5.2 shows the metagenomic summary division for all the four samples analyzed in this study.

Figure 5.3 shows the rarefaction curve of annotated species richness in virus and phage enriched DNA samples. This curve is a plot of the total number of distinct species annotations as a function of the number of sequences sampled. These rarefaction curves are calculated from the table of species abundance. These curves generally rise very quickly at first and then level off towards an asymptote as fewer new species are found per unit of individuals collected. A steeper slope on the left side of the curve indicates that a large fraction of the species diversity remains to be discovered. If the curve becomes flatter to the right, a reasonable number of individuals are sampled: more intensive sampling is likely to yield only few additional species (Meyer et al. 2008, Wilke et al. 2014). The alpha diversity for each sample is also presented in these graphs. Alpha diversity summarizes the diversity of organisms in a sample with a single number; it can be estimated from the distribution of the species-level annotations (Wilke et al. 2014).

According to figure 5.3, diversity within the virus isolated samples is slightly lower in MBR treatment facility as compared to conventional activated sludge (CAS) process. The MBR facility had lower sludge retention time (SRT) compared to conventional WWTP during the month of sampling.

3.2. Virus and phage diversity in sludge:

In order to analyze the diversity of viruses and phages present in sludge samples from a full-scale conventional activated sludge wastewater treatment plant and a membrane bioreactors (MBRs) utility MGRAST v3.0 pipeline was used. Figure 5.4 shows the virus abundance of the all the assembled datasets from the wastewater viromes obtained from MGRAST server based on best hit classification on a family level. The best hit classification option was used which reports the functional and taxonomic annotation of the best hit in the M5NR database for each feature. The number of hits is defined as “occurrences of the input sequence in the database”. *Phycodnaviridae* is the family of viruses present in high numbers (54% and 66% in EL and TC respectively) in sludge samples in both the WWTPs. The *Phycodnaviridae* family consists of large double-stranded-DNA (dsDNA) viruses infecting eukaryotic algae and they are some of the largest known viruses, and have great ecological importance (Larsen et al. 2008). Evidence suggests that these viruses are active players in the formation and termination of algal blooms (Larsen et al. 2008). In EL sludge sample; this family is followed by *Herpesviridae* (10%), *Mimiviridae* (10%) and *Lipothrixviridae* (10%) that are also present in high numbers. Further exploring the *Phycodnaviridae* family on a genus level, *Cholorvirus* is present in higher percentage (57%) followed by *Prasinovirus* (29%) and *Phaeovirus* (14%) (Figure 5.5). Among other virus families, groups of viruses on a genus-level present in EL sludge include

Chlamydia microvirus (50%), *Alphalipothrixvirus* (17%), *Rhadinovirus* (17%), and *Mimivirus* (16%).

In TC sludge samples, the dominating family of viruses following the *Phycodnaviridae* is *Mimiviridae* (16%). Other virus families present are *Asfarviridae* (6%), *Lipothrixviridae* (3%), *Nimaviridae* (3%), *Parvoviridae* (3%), and *Alloherpesviridae* (3%). Further classifications on a genus level, within the *Phycodnaviridae* family, 83% of the viruses are *Chlorovirus* and 17% are *Prasinovirus*. Among other virus families, groups of viruses on a genus-level present in TC sludge include *Mimivirus* (50%), *Asfivirus* (20%), *Betalipothrixvirus* (10%), and *Whispovirus* (10%) and *Parvovirus* (10%). A vast difference in viral diversity at genus level is observed between the two treatment plants. Most of the sequence reads were unclassified that must be derived from putatively novel viruses, therefore has not been taken into percentage analysis. Further classification of virus-isolated samples from EL and TC sludge on a species level is shown in Table 5.3. It is very important to identify the virus diversity as the lack of knowledge on characteristics of the viral universe and the diversity of viral genomes is a roadblock to understanding important issues, like the origin of emerging pathogens and the extent of gene exchange among viruses (Cantalupo et al. 2011). This work presents the diversity viruses in sludge and indicates that there is a difference in the virus community between different treatment process, thus suggesting that within different WWTPs the virus populations are dynamic.

Phage diversity is shown in Figure 5.5. *Siphoviridae*, *Podoviridae* and *Myoviridae* are the most common group of phages found in both the phage enriched samples from sludge obtained from two different treatment plant. Further analyzing each family on a genus level classification as shown in the figure, the dominating genus were *Lambda-like viruses*, *bpp1-like viruses*, and

T4-like viruses respectively in each of the above families. *Myoviridae*, *Podoviridae*, *Siphoviridae*, are also the most common families of phages found in wastewater samples as reported by other studies (Alhamlan et al. 2013; Colomer-Lluch et al. 2011, Parsley et al. 2010). Parsley and his team followed shotgun library approach to study activated sludge samples finding *Myoviridae* (40.3%), *Siphoviridae* (31.9%), *Podoviridae* (25.6%) and considered unclassified phages (2.2%) (Parsley et al. 2010). Pyrosequencing discovered similar results on dairy manure wastewater lagoons where the majority belongs to the family *Siphoviridae* (dsDNA tailed viruses) (Alhamlan et al. 2013). *Myoviridae* and *Siphoviridae* were also detected when electron microscopy was used on sewage and river water sample (Colomer-Lluch et al. 2011). A link between the presence of siphophages and fecal pollution has been suggested implying that these phages could be used as bioindicators of fecal contamination (Rosario et al., 2009).

Further classification in phage enriched samples is shown in Table 5.4 on a species level. Not much difference in the diversity between the two treatment plants is observed with few exceptions like *Streptomyces phage* is only present in TC sample and *Pseudomonas phage* is more abundant in EL sludge.

It has been suggest that bacteriophage may play a significant role in determining the structure and function of bacterial communities in activated sludge (Barr et al. 2010). Considering the potential that many phages kill bacteria, it has been suggested to be used to eradicate bacteria from water (Alhamlan et al. 2013). More research is needed in this direction to further explore this possibility. Despite this critical role in ecosystem processes, the study of virus/phage diversity has lagged far behind parallel studies of the Bacterial and Eukaryotic kingdoms.

Metagenomic analyses of wastewater samples in this study also revealed that a large proportion of sequences could not be assigned to taxonomic affiliations even at the phylum/class levels. Most sequences showed no sequence relation to any known sequences in the databases and thus are most likely to be derived from novel, uncharacterized viruses or phages. This study provided a bioinformatics approach for identifying viral diversity thus adding the knowledge gap to the number of metagenomic datasets previously available. These results present a unique opportunity to examine the WWTPS with respect to virus and phage diversity; these metagenomes provide important data to make baseline observations that will need to be examined more thoroughly in future studies.

4. Conclusions:

In order to evaluate virus and phage community composition in a conventional and MBR utility, metagenomics coupled with Illumina sequencing was used. It provided a broader outlook of the microbial composition in sludge and effluent samples. The analyses unveiled that the communities are highly diverse.

- Virus diversity: Phycodnaviridae are highly abundant followed by *Herpesviridae*, *Lipothrixviridae* and *Mimiviridae* in sludge samples. Difference in viral diversity on a genus level is observed between different treatment plants.
- *Siphoviridae*, *Podoviridae* and *Myoviridae* are the most common group of phages found in both the phage enriched samples from sludge obtained from two different treatment plant.
- Most of the sequences detected were uncharacterized; indicating a greater number of viral diversity is yet to be discovered.

The results presented here revealed the feasibility of metagenomic approaches to characterize viral communities in complex environmental samples, thus describing the role of WWTP in returning safe and healthy water.

5. Acknowledgement:

We would like to thank the managers of the East Lansing Wastewater Treatment Plant and Traverse City Wastewater Treatment Plant for providing the samples and information needed for this study. We would like to acknowledge bioinformatic support and assistance provided by Bioinformatic team at High Performance Computing Center (HPCC) at Michigan State University. A very special thank to Bioinformatic Research Specialist Dr. Tracy Teal.

APPENDIX

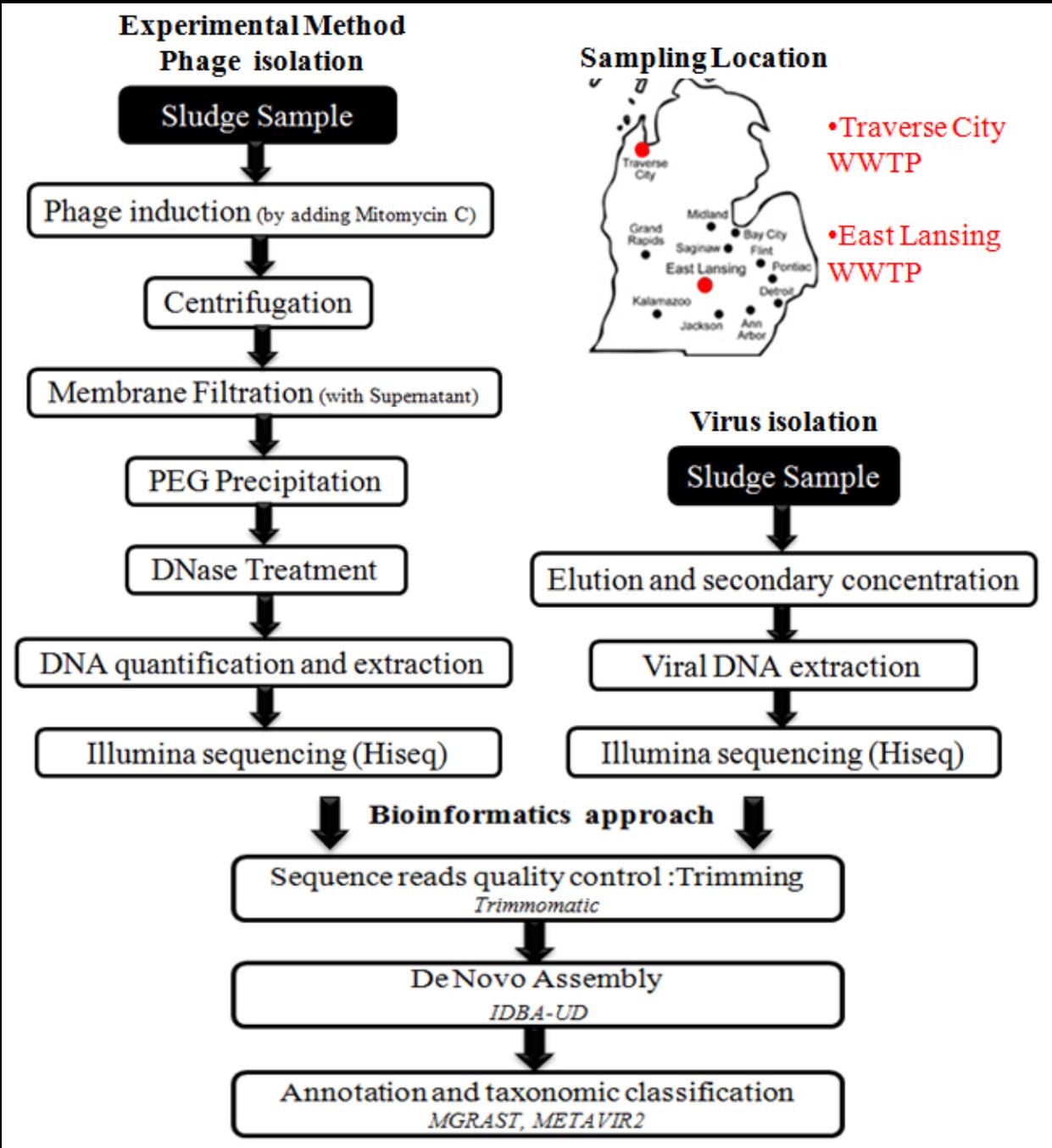


Figure 5.1: Schematic of methodology and location of sampling.

Table 5.1: Characteristics of WWTPs.

	EAST LANSING WWTP	TRAVERSE CITY WWTP
Wastewater treatment process (Biological treatment)	Activated Sludge (AS)	Membrane BioReactor (MBR)
Sludge Retention Time (SRT)	14 days	7.58 days
Capacity	18.8 MGD*	17.0 MGD
Average flow	13.4 MGD	8.5 MGD
Discharge Rate	14.1 MGD	4.0 MGD
<i>Disinfection</i>	<i>Chlorine (Cl)</i>	<i>Ultra-Violet (UV)</i>

Table 5.2: Metagenome analysis statistics for sludge samples (generated by MGRAST).

Metagenome Name	Raw bp count	No. of contigs	Mean Sequence Length (bp)	Post QC: bp Count	Post QC No. of contigs	Post QC: Mean Sequence Length (bp)
EL_VIRUS	187710045	297613	630 ± 652	162566539	289850	560 ± 272
TC_VIRUS	160406503	244963	654 ± 1063	141269383	240752	586 ± 348
EL_PHAGE	118452544	154095	768 ± 1557	90081784	137320	655 ± 551
TC_PHAGE	119319471	138306	862 ± 1763	91302789	125532	727 ± 617

Note: Sequences were assembled and contigs generated were analyzed on MGRAST;

Abbreviation: bp= base pair

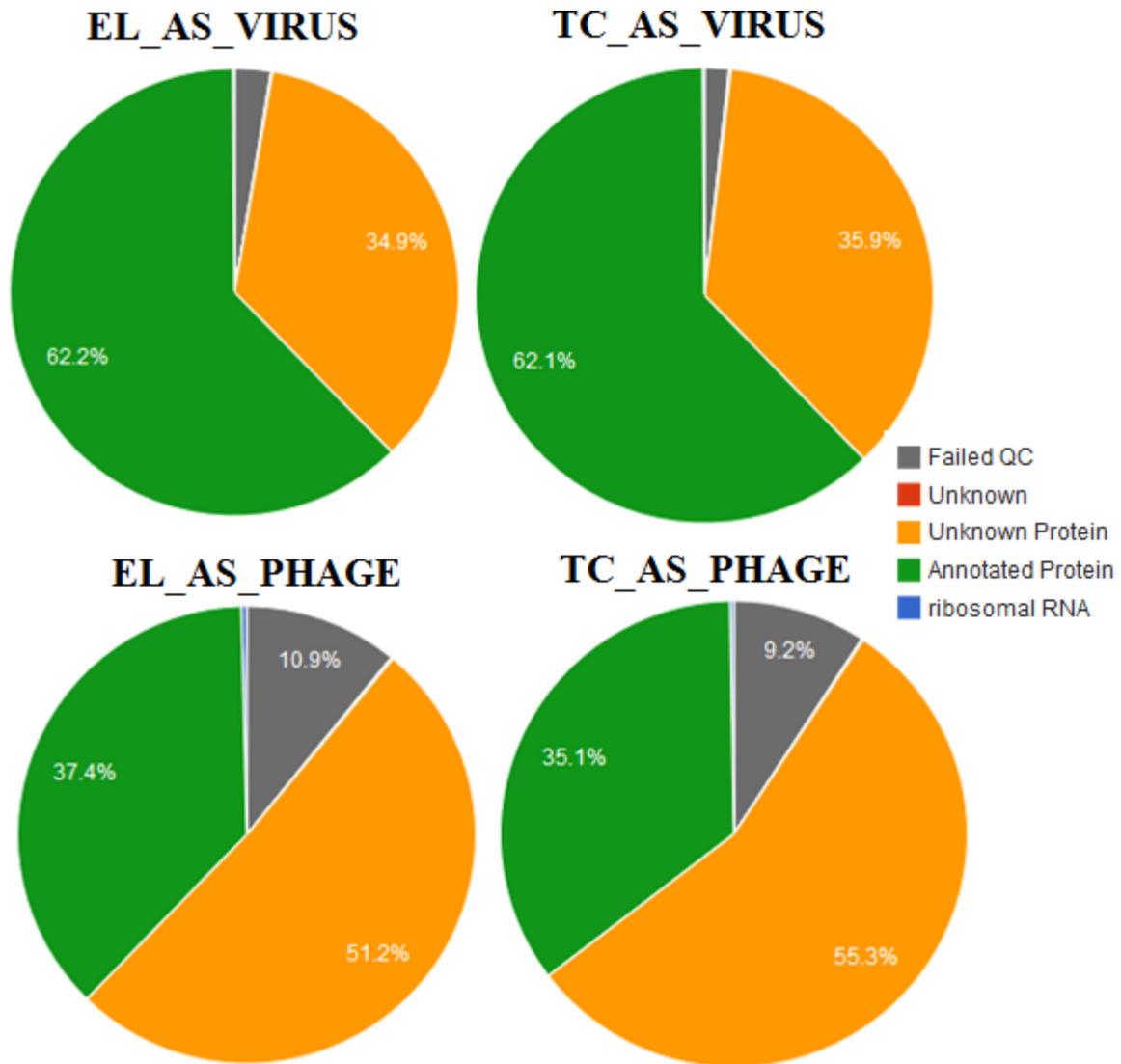
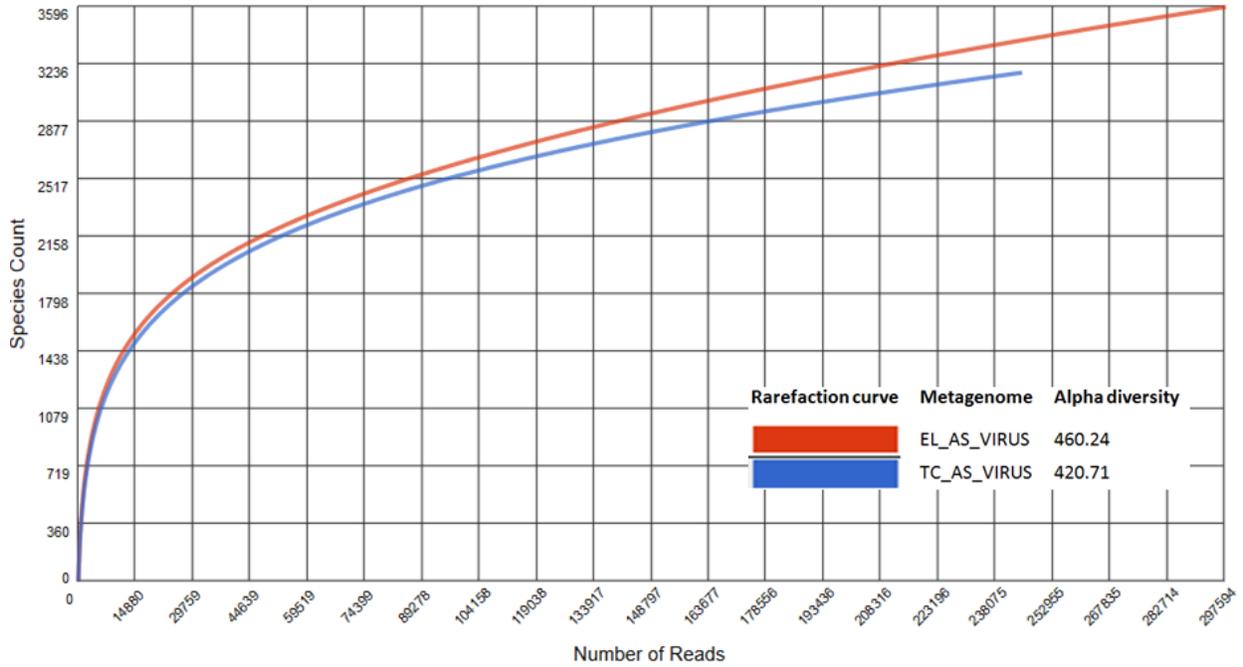
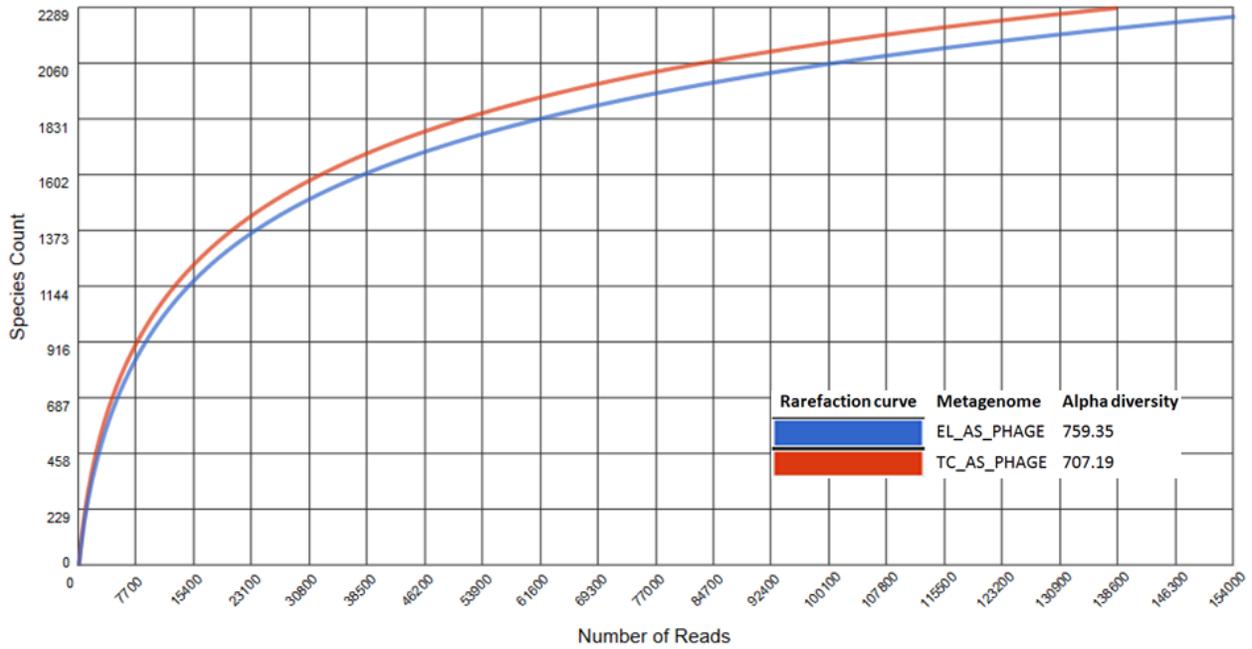


Figure 5.2: Metagenome Summary in phage and virus DNA enriched samples from sludge of two different WWTPs.



(a) Rarefaction curve for viruses in sludge samples



(b) Rarefaction curve for phage in sludge samples

Figure 5.3: Rarefaction curve of species richness in phage and virus DNA enriched samples from sludge of two different WWTPs.

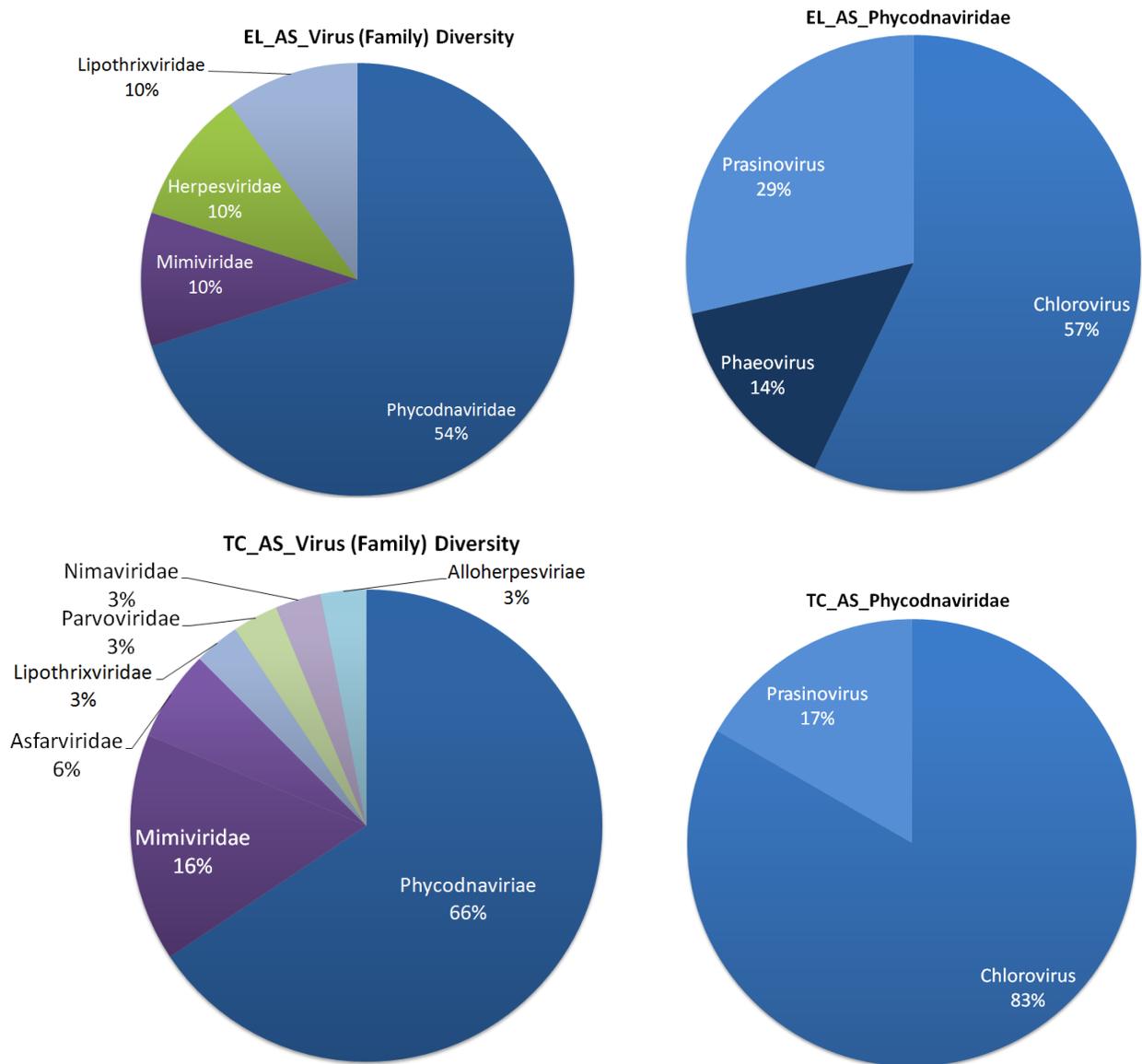


Figure 5.4: Virus diversity/organism abundance (family-level) based on best hit classification: Virus diversity in activated sludge on a family level, further classified into genus level for East Lansing and Traverse City wastewater utilities.

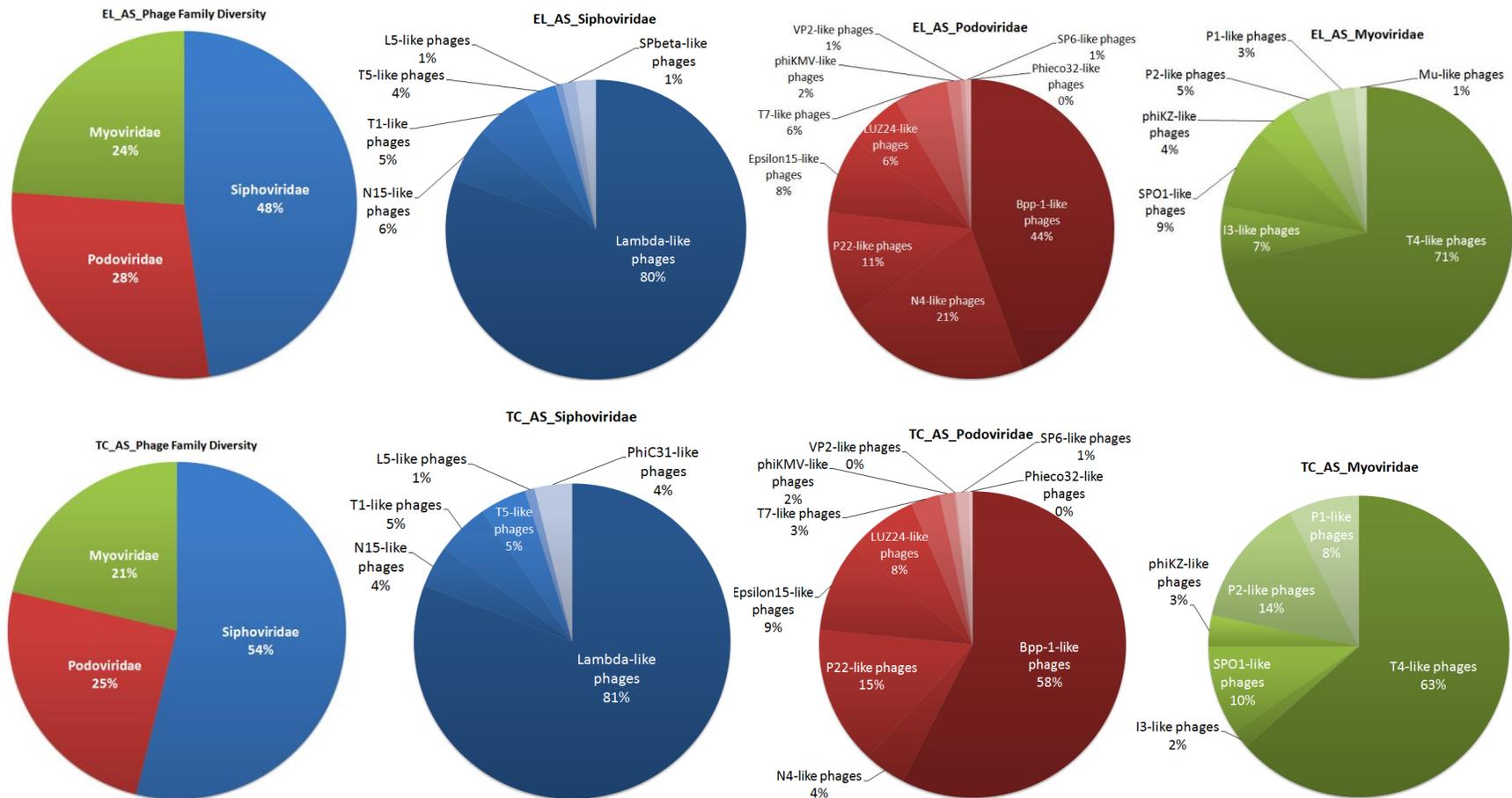


Figure 5.5: Phage diversity/organism abundance based on best hit classification. (a) Diversity of phages on a family level in activated sludge for Traverse City and East Lansing wastewater utilities, (b) Phage diversity (genus-level) within the Siphoviridae, Podoviridae and Myoviridae family for East Lansing and Traverse City wastewater utilities

Table 5.3: Virus abundance in ELWWTP and TCWWTP activated sludge samples.

Virus Abundance				
Family (host)	Genome Group	Genus	East Lansing AS	Traverse City AS
Phycodnaviridae (algae)	dsDNA	Chlorovirus	4	15
		Phaeovirus	1	0
		Prasinovirus	2	3
		unclassified (derived from Phycodnaviridae)	0	3
Mimiviridae (<i>acanthamoeba</i>)	dsDNA	Mimivirus	1	5
Asfarviridae (animal)	dsDNA	Asfivirus	0	2
Lipothrixviridae (archaea)	dsDNA	Alphalipothrixvirus	1	0
		Betalipothrixvirus	0	1
Nimaviridae (animal)	dsDNA	Whispovirus	0	1
Parvoviridae (vertebrates, Insects)	ssDNA	Parvovirus	0	1
Alloherpesviridae (fish)	dsDNA	unclassified (derived from Alloherpesviridae)	0	1
Herpesviridae (vertebrates)	dsDNA	Rhadinovirus	1	0
unclassified (derived from Viruses)			20	136

Virus Abundance_Mimiviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Mimivirus	<i>Acanthamoeba polyphaga</i> <i>mimivirus</i>	1	5

Virus Abundance_Asfarviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Asfivirus	African swine fever virus	0	2

Table 5.3 (cont'd)

Virus Abundance_Lipothrixviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Alphalipothrixvirus	Thermoproteus tenax virus 1	1	0
Betalipothrixvirus	Acidianus filamentous virus 7	0	1

Virus Abundance_Nimaviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Whispovirus	White spot syndrome virus 1	0	1

Virus Abundance_Parvoviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Parvovirus	Porcine parvovirus	0	1

Virus Abundance_Alloherpesviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
unclassified (derived from Alloherpesviridae)	Anguillid herpesvirus 1	0	1

Virus Abundance_Herpesviridae (Genus/Species) Adundance			
Genus	Species	East Lansing AS	Traverse City AS
Rhadinovirus	Leporid herpesvirus 1	1	0

Table 5.4: Phage abundance in ELWWTP and TCWWTP activated sludge samples.

Phage Abundance			
Family	Genus	East Lansing AS	Traverse City AS
Siphoviridae	unclassified	1769	2106
	Lambda-like phages	111	167
	N15-like phages	8	9
	T1-like phages	8	11
	T5-like phages	5	10
	L5-like phages	1	2
	SPbeta-like phages	2	0
	PhiC31-like phages	3	8
Podoviridae	unclassified	558	546
	Bpp-1-like phages	257	299
	N4-like phages	123	23
	P22-like phages	67	76
	Epsilon15-like phages	48	49
	LUZ24-like phages	36	39
	T7-like phages	35	16
	phiKMV-like phages	9	8
	VP2-like phages	3	1
	SP6-like phages	3	7
	Phieco32-like phages	1	2
Myoviridae	unclassified	744	708
	T4-like phages	152	132
	I3-like phages	14	4
	SPO1-like phages	19	20

Table 5.4 (cont'd)

	phiKZ-like phages	9	7
	P2-like phages	10	29
	P1-like phages	6	16
	Mu-like phages	3	0

Phage Abundance_Siphoviridae (Genus/Species)			
(Family) Genus	Species	East Lansing AS	Traverse City AS
Lambda-like phages	Bacillus phage SPP1	2	1
	Burkholderia phage KS9	2	16
	Burkholderia phage phi1026b	16	11
	Burkholderia phage phi644-2	4	7
	Burkholderia phage phiE125	67	70
	Enterobacteria phage HK022	4	4
	Enterobacteria phage HK97	0	18
	Enterobacteria phage cdtI	2	0
	Escherichia Stx1 converting bacteriophage	12	21
	Pseudomonas phage DMS3	0	1
	Streptomyces phage VWB	0	20
	Stx2 converting phage II	0	1
Stx2-converting phage 86	2	0	
T1-like phages	Enterobacteria phage RTP	3	4
	Enterobacteria phage T1	5	2
	Enterobacteria phage TLS	0	5
T5-like phages	Enterobacteria phage SPC35	0	1
	Enterobacteria phage T5	5	9
N15-like phages	Enterobacteria phage N15	8	9

Table 5.4 (cont'd)

PhiC31-like phages	Streptomyces phage phiBT1	0	3
	Streptomyces phage phiSASD1	3	5
L5-like phages	Mycobacterium phage Bxb1	1	0
	Mycobacterium phage D29	0	2
SPbeta-like phages	Bacillus phage SPbeta	2	0

Phage Abundance_Podoviridae (Genus/Species)			
(Family) Genus	Species	East Lansing AS	Traverse City AS
Bpp-1-like phages	Bordetella phage BIP-1	178	167
	Bordetella phage BMP-1	173	166
	Bordetella phage BPP-1	186	184
	Burkholderia phage BcepC6B	66	114
N4-like phages	Enterobacter phage EcP1	27	7
	Enterobacteria phage N4	39	9
	Pseudomonas phage LIT1	10	1
	Pseudomonas phage LUZ7	32	2
	Roseovarius sp. 217 phage 1	0	1
	Silicibacter phage DSS3phi2	8	3
	Sulfitobacter phage EE36phi1	7	0
P22-like phages	Enterobacteria phage CUS-3	14	17
	Enterobacteria phage P22	3	13
	Enterobacteria phage ST104	19	22
	Myxococcus phage Mx9	1	2
	Salmonella phage HK620	10	5
	Salmonella phage SE1	1	1
	Salmonella phage ST160	5	7
	Salmonella phage ST64T	11	8

Table 5.4 (cont'd)

	Salmonella phage epsilon34	1	0
	Salmonella phage g341c	5	2
	Shigella phage Sf6	0	3
Epsilon15-like phages	Escherichia phage phiV10	19	31
	Salmonella phage epsilon15	29	18
LUZ24-like phages	Pseudomonas phage LUZ24	24	25
	Pseudomonas phage PaP3	12	14
T7-like phages	Enterobacteria phage 13a	0	3
	Enterobacteria phage BA14	3	5
	Enterobacteria phage EcoDS1	5	4
	Enterobacteria phage K1F	0	1
	Enterobacteria phage T3	3	0
	Enterobacteria phage T7	4	0
	Klebsiella phage K11	3	0
	Kluyvera phage Kvp1	1	0
	Pseudomonas phage gh-1	2	0
	Pseudomonas phage phiIBB-PF7A	1	0
	Salmonella phage phiSG-JL2	1	0
	Vibriophage VP4	1	0
	Yersinia pestis phage phiA1122	4	0
	Yersinia phage Berlin	0	1
	Yersinia phage Yepe2	1	1
	Yersinia phage phiYeO3-12	6	1
phiKMV-like phages	Pseudomonas phage LKA1	0	2
	Ralstonia phage RSB1	4	4
	Vibrio phage VP93	5	2
VP2-like phages	Vibrio phage VP2	2	1

Table 5.4 (cont'd)

	Vibrio phage VP5	3	1
SP6-like phages	Enterobacteria phage Era103	3	6
	Enterobacteria phage K1-5	0	1
	Erwinia phage phiEa100	3	4
	Erwinia phage phiEa1H	3	4
Phieco32-like phages	Enterobacteria phage Phieco32	1	2

Phage Abundance_Myoviridae (Genus/Species)			
(Family) Genus	Species	East Lansing AS	Traverse City AS
T4-like phages	Acinetobacter phage 133	4	4
	Acinetobacter phage Ac42	2	8
	Acinetobacter phage Acj61	3	5
	Acinetobacter phage Acj9	2	5
	Aeromonas phage 25	2	2
	Aeromonas phage 31	1	0
	Aeromonas phage 44RR2.8t	1	1
	Aeromonas phage 65	7	15
	Aeromonas phage Aeh1	9	0
	Aeromonas phage PX29	1	0
	Aeromonas phage phiAS4	1	0
	Aeromonas phage phiAS5	7	6
	Enterobacteria phage JSE	3	0
	Enterobacteria phage Phi1	0	1
	Enterobacteria phage RB32	1	0
	Enterobacteria phage RB43	4	1
	Enterobacteria phage RB49	5	0
	Enterobacteria phage RB69	0	1

Table 5.4 (cont'd)

	Enterobacteria phage T4 sensu lato	5	5
	Enterobacteria phage vB_EcoM-VR7	9	4
	Klebsiella phage KP15	0	1
	Prochlorococcus phage P-SSM2	22	13
	Prochlorococcus phage P-SSM4	30	25
	Synechococcus phage S-PM2	11	12
	Synechococcus phage syn9	19	13
	Vibrio phage nt-1 sensu lato	3	10
P2-like phages	Aeromonas phage phiO18P	0	3
	Enterobacteria phage P2	1	4
	Pseudomonas phage phiCTX	0	1
	Ralstonia phage phiRSA1	9	21
SPO1-like phages	Bacillus phage Bastille	3	4
	Bacillus phage SPO1	4	0
	Lactobacillus phage LP65	4	6
	Listeria phage A511	2	6
	Staphylococcus phage G1	4	1
	Staphylococcus phage K	0	1
	Staphylococcus phage Twort	2	3
P1-like phages	Enterobacteria phage P1	6	16
I3-like phages	Mycobacterium phage Bxz1	7	1
	Mycobacterium phage Cali	13	1
	Mycobacterium phage Catera	1	1
	Mycobacterium phage ET08	8	1
	Mycobacterium phage Rizal	7	3
	Mycobacterium phage Spud	7	1

Table 5.4 (cont'd)

phiKZ-like phages	Pseudomonas phage 201phi2-1	5	5
	Pseudomonas phage EL	4	0
	Pseudomonas phage phiKZ	0	2
Mu-like phages	Burkholderia phage BcepMu	3	0

REFERENCES

REFERENCES

1. Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R., & Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *Journal of microbiological methods*, 92(2), 183-188.
2. Aw, T. G., Howe, A., & Rose, J. B. (2014). Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *Journal of virological methods*.
3. Barr, J. J., Slater, F. R., Fukushima, T., & Bond, P. L. (2010). Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS microbiology ecology*, 74(3), 631-642.
4. Bergh, Ø., Børshem, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, 340(6233), 467-468.
5. Bibby, K. (2014). Improved Bacteriophage Genome Data is Necessary for Integrating Viral and Bacterial Ecology. *Microbial ecology*, 67(2), 242-244.
6. Bibby, K., & Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental science & technology*, 47(4), 1945-1951.
7. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170
8. Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., ... & Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *MBio*, 2(5), e00180-11.
9. Carter, J., & Saunders, V. A. (2007). *Virology: principles and applications*. John Wiley & Sons.
10. Colomer-Lluch, M., Jofre, J. and Muniesa, M. (2011) Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One* 6(3), e17549.
11. Ewert, D.L. and Paynter, M. (1980) Enumeration of bacteriophages and host bacteria in sewage and the activated-sludge treatment process. *Applied and environmental microbiology* 39(3), 576-583.
12. Gomila, M., Solis, J., David, Z., Ramon, C., & Lalucat, J. (2008). Comparative reductions of bacterial indicators, bacteriophage-infecting enteric bacteria and enteroviruses in wastewater tertiary treatments by lagooning and UV-radiation.
13. Havelaar, A., Pot-Hogbeem, W., Furuse, K., Pot, R. and Hormann, M. (1990) F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. *Journal of Applied Microbiology* 69(1), 30-37.

14. Khan, M.A., Satoh, H., Katayama, H., Kurisu, F. and Mino, T. (2002) Bacteriophages isolated from activated sludge processes and their polyvalency. *Water research* 36(13), 3364-3370.
15. Mandilara, G.D., Smeti, E.M., Mavridou, A.T., Lambiri, M.P., Vatopoulos, A.C. and Rigas, F.P. (2006) Correlation between bacterial indicators and bacteriophages in sewage and sludge. *FEMS microbiology letters* 263(1), 119-126.
16. Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R. and Wilke, A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* 9(1), 386.
17. Muniesa, M., Blanco, J.E., de Simón, M., Serra-Moreno, R., Blanch, A.R. and Jofre, J. (2004b) Diversity of stx2 converting bacteriophages induced from Shiga-toxin-producing *Escherichia coli* strains isolated from cattle. *Microbiology* 150(9), 2959-2971.
18. Otawa, K., Lee, S.H., Yamazoe, A., Onuki, M., Satoh, H. and Mino, T. (2007) Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes. *Microbial ecology* 53(1), 143-152.
19. Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper, W.F., Jr. and Liles, M.R. (2010) Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl Environ Microbiol* 76(11), 3753-3757.
20. Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420-1428.
21. Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., & Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental microbiology*, 11(11), 2806-2820.
22. Rosenberg, E., Bittan-Banin, G., Sharon, G., Shon, A., Hershko, G., Levy, I. and Ron, E.Z. (2010) The phage-driven microbial loop in petroleum bioremediation. *Microbial Biotechnology* 3(4), 467-472.
23. Sander, M. and Schmieger, H. (2001) Method for host-independent detection of generalized transducing bacteriophages in natural habitats. *Appl Environ Microbiol* 67(4), 1490-1493.
24. Shapiro, O.H. and Kushmaro, A. (2011). Bacteriophage ecology in environmental biotechnology processes. *Curr Opin Biotechnol* 22(3), 449-455.
25. Sidhu, J., Hanna, J., & Toze, S. (2008). Survival of enteric microorganisms on grass surfaces irrigated with treated effluent. *Journal of water and health*, 6(2), 255-262.
26. Toze, S. (1997). *Microbial Pathogens in Wastewater: Literature review for urban water systems multi-divisional research program*. CSIRO Land and Water.

27. USEPA, 2001. Manual of Methods for Virology (Chapter 14). EPA 600/4-84/013 Office of Water, U.S. Environmental Protection Agency, Washington, DC.
28. Wilke, A., Glass, E. M., Bischof, J., Braithwaite, D., D'Souza, M., Gerlach, W., ... & Meyer, F. (2014). MG-RAST Manual for version 3.3. 6, revision 6.
29. Withey, S., Cartmell, E., Avery, L. M., & Stephenson, T. (2005). Bacteriophages—potential for application in wastewater treatment processes. *Science of the total environment*, 339(1), 1-18.
30. Wu, Q. and Liu, W.-T. (2009). Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Water research* 43(4), 1101-1109.
31. Xagorarakis, I., Yin, Z., & Svambayev, Z. (2014). Fate of Viruses in Water Systems. *Journal of Environmental Engineering*.
32. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., & Treiber, G. (1970). Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology*, 40(3), 734-744.

CHAPTER 6

CONCLUSIONS AND SIGNIFICANCE

Our environment is greatly impacted by the presence of emerging microbial contaminants. The overall objective of this study was to provide metagenomic insights into bacterial, viral and phage diversity and resistance to antibiotics and metal compounds in wastewater utilities.

Chapter 1 focuses on phage diversity and antibiotic resistance genes in a conventional wastewater treatment plant in Michigan. A method for phage DNA isolation was optimized using PEG (polyethylene glycol) precipitation and DNase (deoxyribonuclease) treatment. Phage DNA was screened for ARGs (tetracycline resistant genes (Tet-W and Tet-O) and sulfonamide resistant gene (Sul-I)) using real-time Q-PCR. Diversity of phages was studied by next generation sequencing with Illumina (Miseq). Phage metagenomes were searched for functional signatures of antibiotic resistance genes. Metagenomics analysis revealed that most of the observed antibiotic resistance was resistance to methicillin, fluoroquinolones and beta-lactamase group of antibiotics. The findings of the study also suggest that there is a substantial shift in the phage community over the course of the activated sludge process between primary and returned activated sludge, thus suggesting that within the activated sludge the phage populations are dynamic and that phage DNA was associated with antibiotic resistant genes in wastewater.

Chapter 2 focuses on the diversity of microbial resistances to antibiotics and metal compounds in samples from a conventional and an MBR (membrane biological Reactor) utility using metagenomic investigations. Illumina Hiseq sequencing was applied on six samples from the two different WWTPs in Michigan where bacterial DNA was isolated from three different

sampling points (activated sludge, before disinfection effluent and after disinfection effluent). Findings of this study reveal that genes coding for antibiotic resistance were identified in all bacterial samples, along with genes coding for resistance to metals. The MBR utility showed slightly higher number of hits for all the functional categories compared to conventional WWTP samples. The incidence of multiple metal and antibiotic resistances among bacterial populations poses a potential threat to human health.

Chapter 3 focuses on identification of viral and bacterial pathogen diversity in wastewater effluent released from an MBR and a conventional treatment utility using metagenomics analysis. Effluent samples (before and after disinfection) have been collected and analyzed to reveal microbial pathogenic diversity. Findings of this study show that a potential human viral pathogen observed in our samples belongs to taxonomic order *Herpesviridales*. Other potentially pathogenic viruses detected in this study include Adenoviridae, Coronaviridae and Hepatitis C viruses. Whereas all the bacterial pathogens described in the Contaminant Candidate List from EPA were detected in our study. Diversity analysis does not provide quantitative data on pathogen loads or infectivity but it provides a list of potentially pathogenic viruses and bacteria that need to be considered in more detail.

Chapter 4 focuses on investigating the phage and virus diversity of two distinct wastewater treatment systems, conventional and MBR using metagenomics. Samples collected from activated sludge were studied using Illumina Hiseq sequencing. Analysis software MGRAST v3.0 pipeline was used for analyzing the assembled sequences. Our results showed differences in viral community compositions on a genus level between two different WWTPs, suggesting there is a considerable difference in the community between different treatment

processes. This study provided a bioinformatics approach for identifying viral diversity which was an important gap in knowledge.

The work presented in chapter 1-4 is significant since it characterizes the microbial diversity in wastewater. Increasing population and urbanization will result into increasing wastewater quantities. Management of ARGs and microbial pathogens in WWTPs will be critical since the problem of spread of pathogens and antibiotic resistant genes in the environment poses a significant challenge to public health management. Full-scale conventional and state of the art wastewater utilities have been found to release pathogens and resistant bacteria in the environment. Understanding the microbial diversity and linkage between diversity and wastewater treatment methods and practices will lead to sustainable wastewater management. More research is needed to study links between engineering design, microbial diversity and operational parameters.